

Diseño y desarrollo de tesauros

Ana María Martínez Tamayo - Paola Verónica Mendes



Diseño y desarrollo de tesauros

Ana M. Martínez Tamayo

Paola V. Mendes

Facultad de Humanidades y Ciencias de la Educación
Universidad Nacional de La Plata

2015

Esta publicación ha sido sometida a evaluación interna y externa organizada por la Secretaría de Investigación de la Facultad de Humanidades y Ciencias de la Educación de la Universidad Nacional de La Plata.

Diseño: D.C.V. Federico Banzato

Tapa: D.G. P. Daniela Nuesch

Asesoramiento imagen institucional: Área de Comunicación Visual

Corrección: Lic. Alicia Lorenzo

Queda hecho el depósito que marca la ley 11.723

©2015 Universidad Nacional de La Plata.

ISBN 978-950-34-1307-4

Colección Biblioteca Humanidades 38

Cita sugerida: Martínez Tamayo, A. M. y Mendes. P. V. (2015). Diseño y desarrollo de tesauros. La Plata : Universidad Nacional de La Plata. Facultad de Humanidades y Ciencias de la Educación. (Biblioteca Humanidades ; 38)

Recuperado de <http://www.libros.fahce.unlp.edu.ar/index.php/libros/catalog/book/68>



Licencia Creative Commons 2.5 a menos que se indique lo contrario

Universidad Nacional de La Plata
Facultad de Humanidades y Ciencias de la Educación

Decano

Dr. Aníbal Viguera

Vicedecano

Dr. Mauricio Chama

Secretaria de Asuntos Académicos

Prof. Hernán Sorgentini

Secretario de Posgrado

Dr. Fabio Espósito

Secretaria de Investigación

Dra. Susana Ortale

Secretario de Extensión Universitaria

Mg. Jerónimo Pinedo

Índice

Etiquetas y símbolos usados en el tesoro	6
Abreviaturas y siglas	7
Prefacio	10
La organización del conocimiento	12
Sistemas y procesos de organización del conocimiento	32
Distintos enfoques para estudiar los SOC	49
El tesoro	65
La historia del tesoro	75
Diseño y desarrollo del tesoro	86
El vocabulario del tesoro	99
Las relaciones jerárquicas y asociativas	123
El orden sistemático del tesoro	134
Evaluación, publicación y actualización del tesoro	153
El tesoro y la interoperabilidad	179
Software de gestión, modelo de datos y formatos de intercambio	194
Bibliografía	202
Las Autoras	221
Índice de figuras y tablas	223

Etiquetas y símbolos usados en el tesoro

A continuación, se muestran las etiquetas y símbolos utilizados para las notas, notación y relaciones. Están basados en la norma ISO 25975-1:2005, pero debe tenerse en cuenta que las recomendaciones de la norma para el español no están completas, por lo que hemos agregado de nuestra parte las etiquetas NH, TGE y TEE.

Descripción	Símbolo	Etiqueta	Significado
Elementos descriptivos		NA	Nota de alcance
		DEF	Nota de definición
		NH	Nota histórica
Elementos clasificatorios		CC	Notación
		CM	Categoría de materia
Relaciones	→	USE	Remite de un término no preferente a un término preferente.
	=	UP	Informa los términos no preferentes que han sido descartados
		USE ... +	Términos que deben poscoordinarse en sustitución de un término multipalabra
		UP+	El término que sigue es un término compuesto no preferente
		TT	Término tope
	<	TG	Término genérico
		TGG	Término genérico genérico
	<C	TGP	Término genérico partitivo
		TGE	Término genérico enumerativo
	>	TE	Término específico
		TEG	Término específico genérico
	>C	TEP	Término específico partitivo
		TEE	Término específico enumerativo
	-	TR	Término relacionado
Interoperabilidad		EQ	Equivalencia entre términos de dos o más tesauros
		+EQ	Equivalencia exacta entre términos de dos o más tesauros
		-EQ	Equivalencia inexacta
		EQ ... +	Equivalencia compuesta
		EQ [...]	Equivalencia compuesta por intersección
		BM	Maapeo genérico (broader mapping)
		NM	Maapeo específico (narrower mapping)
		RM	Maapeo relacionado (related mapping)

Abreviaturas y siglas

AENOR: Asociación Española de Normalización y Certificación
ANSI: American National Standard Institute
ARMARC: Archivo de Autoridades de Materia en Formato MARC
ASC: Afrika Studiecentrum
ASTIA: Armed Services Technical Information Abstracts
BC: Bibliographic classification
BMD: Bureau Marcel van Dijk
BNF: Bibliothèque Nationale de France
BS: British standard [norma británica]
BSI: British Standards Institute
CAICYT: Centro Argentino de Información Científica y Técnica
CAT: Chinese Agricultural Thesaurus
CC: Colon classification
CDU: Clasificación decimal universal
COSATI: Committee on Scientific and Technical Information
CRG: Classification Research Group
DCMI: Dublin Core Metadata Initiative
DeCS: Descriptores en Ciencias de la Salud
DOI: Digital Object Identifier
FRAD: Requisitos funcionales para datos de autoridad
FRBR: Requisitos funcionales para registros bibliográficos
FRSAD: Requisitos funcionales para datos de autoridad de materia
ICONTEC: Instituto Colombiano de Normas Técnicas y Certificación
IFLA: International Federation of Library Associations and Institutions
IRAM: Instituto Argentino de Normalización
ISBN: Número estándar internacional de libros
ISKO: International Society for Knowledge Organization

ISO: International Organization for Standardization
ISSN: International Standard Serial Number
IUBMB: International Union of Biochemistry and Molecular Biology
IUPAC: International Union of Pure and Applied Chemistry
IZ: Thesaurus Socialwissenschaften
KB: Koninklijke Bibliotheek
KWAC: Keyword and context (palabra clave y contexto)
KWIC: Keyword in context (palabra clave en el contexto)
KWOC: Keyword out of context (palabra clave fuera de contexto)
LCC: Library of Congress classification
LCSH: Library of Congress subject headings
LEMB: Lista de encabezamientos de materia para bibliotecas
MADS: Metadata authority description schema
MARC Machine Readable Cataloging
MeSH: Medical subject headings
NISO: National Information Standards Organization
OAI-PMH: Open Archives Initiative Protocol Metadata Harvesting
OC: Organización del conocimiento
OIT: Organización Internacional del Trabajo
ONU: Organización de las Naciones Unidas
OWL: Web ontology language
PMEST: Personalidad, materia, energía, espacio y tiempo
POC: Procesos de organización del conocimiento
RDF: Resource description framework
SAR: Sistema de almacenamiento y recuperación
SCDD: Sistema de clasificación decimal de Dewey
Sears: Lista de encabezamientos de materia para bibliotecas pequeñas de Sears
SKOS: Simple knowledge organization system (Sistemas simples de organización del conocimiento)
SOC: Sistema de organización del conocimiento
SWD: Schlagwortnormdatei
TEE: Tesauro europeo de educación
UAT: Unified astronomy thesaurus
UML: Unified modelling language

UNBIS: United Nations Bibliographic Information System

UNESCO: Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura

URI: Uniform resource identifier

W3C: World Wide Web Consortium

Web: World Wide Web

WWW: World Wide Web

XML: Extensible markup language

Prefacio

La presente obra, de carácter introductorio, está dirigida a estudiantes y profesionales de bibliotecología, documentación o ciencia de la información de lengua española que se enfrentan por primera vez al tema del diseño y desarrollo de tesauros.

Una objeción que puede hacerse a este texto es que ya se han publicado otros escritos de reconocidos autores sobre este tema. También existen diversas normas de organismos normalizadores y especializados en esta área, particularmente la International Organization for Standardization (ISO) y la International Federation of Library Associations and Institutions (IFLA). Varios de estos manuales y de estas normas se citan aquí y nos reconocemos en deuda con todos ellos.

Sin embargo, como afirma un viejo refrán, “cada maestrillo con su librillo”. Nos importa encuadrar el diseño y desarrollo de tesauros dentro del marco referencial de la organización del conocimiento, entendida en su relación con la bibliotecología, la documentación y la ciencia de la información, tal como ha sido recomendado por el IV Encuentro de Directores de Escuelas de Bibliotecología y III Encuentro de Docentes de Bibliotecología del Mercosur, celebrado en Montevideo en 2000. Este marco referencial nos obliga a considerar el tesoro como un sistema de organización del conocimiento y a usar una terminología acorde.

Por otra parte, y a diferencia de los textos citados en esta obra, recogemos el aporte de las nuevas normas de diseño y desarrollo de tesauros publicadas en este siglo veintiuno, especialmente la norma internacional ISO 25964-1/2, que ha puesto el énfasis en la publicación electrónica del tesoro y en el aprovechamiento de los recursos informáticos para la interoperabilidad entre los tesauros y otros sistemas de organización del conocimiento.

Por supuesto, también nos interesa reflejar nuestros puntos de vista, así

como las coincidencias y las diferencias con otros autores. Nada de esto va en detrimento de las distintas ideas o posiciones.

En los capítulos 1 y 2 hacemos una introducción a la organización del conocimiento, así como a sus sistemas y procesos. En el capítulo 3 se analizan distintos enfoques con los que se han estudiado los sistemas de organización del conocimiento. Estos tres capítulos constituyen el marco de referencia en el que insertamos el tesoro y su diseño y desarrollo.

Adentrándonos en el tema específico de esta obra, en los capítulos 4 y 5 ofrecemos una definición de tesoro y su caracterización, así como una breve reseña de su origen y su razón de ser.

En el capítulo 6 presentamos el diseño y desarrollo de tesoros como una cadena de procesos de organización del conocimiento, basándonos en conocidas técnicas de gestión. Un apartado especial en este capítulo está dedicado al principio de garantía en el que deben basarse todas las decisiones del diseño y desarrollo de un tesoro.

En el capítulo 7 se profundiza sobre el vocabulario del tesoro, sus formas gramaticales, la desambiguación y la relación de equivalencia. A continuación, en el capítulo 8 se procede a explicar las relaciones jerárquicas y asociativas, mientras que en el capítulo 9 se describe el orden sistemático del tesoro, que basamos en el análisis de facetas.

En el capítulo 10 nos referimos a la evaluación, publicación y actualización como procesos finales y continuos del diseño y desarrollo del tesoro. En el capítulo 11 abordamos los modelos, métodos, experiencias y dificultades de la interoperabilidad entre el tesoro y otros sistemas de organización del conocimiento. Por último, en el capítulo 12 resumimos las recomendaciones para el programa informático de gestión de tesoros, el modelo de datos y los formatos de intercambio propuestos para su diseño y desarrollo.

Finalmente, queremos manifestar nuestro agradecimiento a la Facultad de Humanidades y Ciencias de la Educación de la Universidad Nacional de La Plata, en especial a la Prosecretaría de Gestión Editorial y Difusión, por la posibilidad que nos han brindado de publicar el presente texto.

Es nuestro deseo que esta obra sea útil para aquellas personas a quienes está dirigida. También deseamos que pueda enriquecerse con las críticas, comentarios y sugerencias que nos hagan conocer.

Las autoras

La organización del conocimiento

En esta obra, entendemos que el tesoro es un sistema de organización del conocimiento y que su diseño y desarrollo constituyen un conjunto de procesos de organización del conocimiento. Para establecer el marco de referencia pertinente, comenzaremos por revisar las distintas concepciones de la organización del conocimiento -una disciplina emergente- y la diferenciaremos de otras áreas.

El término *organización del conocimiento*

La organización del conocimiento -en el sentido que aquí se trata- tiene una historia varias veces milenaria, que puede remontarse, entre otros antecedentes, a las bibliotecas sumerias dos mil años a. C. (Dalby, 1986). En tiempos más cercanos, la organización del conocimiento evolucionó dentro de la bibliotecología, la documentación y la ciencia de la información (en adelante BDCI), con los aportes de figuras como los estadounidenses Melvil Dewey, Charles Ammy Cutter, Henry Evelyn Bliss, Mortimer Taube, Calvin Moers, Hans Peter Luhn, los belgas Paul Otlet y Henri La Fontaine, el bibliotecario indio Shiyali Ramamrita Ranganathan y los miembros del Classification Research Group (CRG) de Gran Bretaña, entre los cuales podemos mencionar a Bernard I. Palmer, Arthur J. Wells, Brian C. Vickery, Douglas J. Foskett, Barbara Kyle, Jack Mills, Derek W. Langridge, Eric J. Coates, Jason Farradane y Derek Austin.

Se acepta que el término *organización del conocimiento* fue acuñado por Henry E. Bliss en dos de sus obras: *The organization of knowledge and the system of the science* y *The organization of knowledge in libraries and the subject approach to books*. Para Bliss, la organización o clasificación de los libros en las bibliotecas debía ser consistente con el desarrollo de la ciencia, es decir, del conocimiento. Este conocimiento -expresado en la investiga-

ción, la enseñanza y la sistematización de los estudios en las universidades, así como en las publicaciones académicas- debía determinar el orden de los libros en el estante y de los registros bibliográficos en el catálogo sistemático (Bliss, 1929, 1933).

Inicialmente, el término *organización del conocimiento* no tuvo mayor repercusión, pero en la segunda mitad del siglo veinte comenzó a usarse de nuevo a partir de las obras del profesor estadounidense Jesse H. Shera (*Documentation and the organization of knowledge*) en 1966 y de los profesores alemanes Dagobert Soergel (*Dokumentation und Organisation des Wissens*) e Ingetraut Dahlberg (*Grundlagen universaler Wissensordnung*) en 1971 y 1974, respectivamente.

Desde entonces, la organización del conocimiento ha ido ganando espacio como un nuevo campo de estudio, particularmente a partir de 1989. Ese año, la entonces International Society for Classification cambió su nombre por International Society for Knowledge Organization (ISKO) y su revista *International Classification* pasó a llamarse *Knowledge Organization*. La institucionalización de la disciplina y su órgano de difusión no solo han dado nuevo impulso a la idea de Bliss, sino que proporcionan además un fuerte respaldo académico.

La corriente de ISKO ha encontrado eco en la región del Mercosur, donde distintos autores la han enriquecido con sus contribuciones. Más aún, en el IV Encuentro de Directores y III de Docentes de Escuelas de Bibliotecología del Mercosur, celebrado en Montevideo del 24 al 27 de mayo de 2000, se resolvió recomendar a la organización del conocimiento como marco referencial para la enseñanza de los temas que aquí nos ocupan (Arest & Ceretta, 2001).

¿Qué es la organización del conocimiento?

La especialista alemana Ingetraut Dahlberg, fundadora y ex presidente de ISKO, recuerda que la existencia de una ciencia descansa sobre el hecho de que posee su propio objeto de estudio y su propia área de actividad. Así, explica que en el caso de la organización del conocimiento ambos elementos se encuentran presentes en el mismo nombre de la disciplina: el conocimiento (objeto) y la organización (actividad). Por tanto, cualquier declaración científica relacionada con ella debe reducirse a ambos conceptos (Dahlberg, 2006a).

Respecto al objeto, es decir, el *conocimiento*, Dahlberg considera que si bien es cierto que el conocimiento existe en la mente de una persona y, por lo tanto, es subjetivo e individual, también es cierto que puede compartirse gracias a la capacidad de los seres humanos para tratar con las cosas del mundo y a su habilidad lingüística para expresar experiencias y puntos de vista. Entonces, ¿a qué conocimiento se refiere la organización del conocimiento?

Para responder esta pregunta recurrimos al profesor Mario G. Barité de la Universidad de la República (Uruguay):

El objeto de estudio de esta disciplina [organización del conocimiento] es el conocimiento socializado, entendiendo por tal todo nuevo conocimiento que se publica y socializa por distintas vías, la mayoría de las cuales persigue registrar la paternidad y la defensa de los derechos consiguientes de sus autores o creadores, tanto como la efectiva difusión del conocimiento generado o sus aplicaciones (Barité, 2001: 227).

El profesor José Augusto Chaves de Guimarães, de la Universidade Estadual Paulista *Júlio de Mesquita Filho* (Brasil), coincide con lo anterior cuando afirma:

El objeto de investigación de la Organización del Conocimiento reside en aquello que Dahlberg (1993: 214) denomina “conocimiento en acción”, o sea, algo de lo cual existe un cierto consenso social, un conocimiento registrado y socializado, cuya organización y representación serán desarrollados de modo que a partir de él pueda ser generado nuevo conocimiento (Guimarães, 2008: 85).

En el mismo sentido, el consultor británico Alan Gilchrist hace referencia a los profesores japoneses Ikujiro Nonaka e Hirotaka Takeuchi, quienes distinguen entre el conocimiento tácito (almacenado en la mente de un individuo) y el conocimiento explícito (registrado en un documento). Gilchrist (2011) señala que la única aproximación al conocimiento en el área de la BDCI es a través del conocimiento explícito.

Ahora bien, para algunos autores el conocimiento comunicado es justamente la información, como se deduce de las siguientes afirmaciones:

Conocimiento es lo que yo sé. Información es lo que *nosotros* sabemos, es decir, el conocimiento compartido (Foskett, 1996: 3).

Información es “el contenido de un mensaje” o “algo que es comunicado” (Svenonius, 2000: 7).

Información: conocimiento que es comunicado (ISO, 2001: 8).

La comunicación del conocimiento puede ser oral, como sucede por ejemplo cuando se diserta en una conferencia o cuando se debate en una mesa redonda. Para que pueda ser organizado en las bibliotecas, los centros de documentación o la World Wide Web, se requiere que el conocimiento socializado se registre en un soporte físico (papel, registros audiovisuales, memorias informáticas, etc.), con lo cual se convierte en un *documento* (ISO, 2001: 11).

Entonces, el objeto de la organización del conocimiento es el conocimiento socializado y registrado, explícito o documentado.

En cuanto a la actividad *organización*, Dahlberg (2006a) distingue dos aplicaciones:

- Construcción de sistemas de conceptos: consiste en estructurar el conocimiento contenido en las unidades de conocimiento (conceptos) en un orden sistemático, de acuerdo con sus características, para crear un sistema de conceptos, por ejemplo un tesoro o un sistema de clasificación. Otros autores prefieren llamarlos *sistemas de organización del conocimiento* (SOC), término que adoptamos en esta obra.
- Correlación de los conceptos con los objetos de la realidad: se refiere a la correlación de los conceptos contenidos en los SOC con aquellos expresados en los documentos. Esta correlación incluye: a) la asignación de notaciones tomadas de un sistema de clasificación y b) la asignación de términos controlados tomados por ejemplo de un tesoro, o palabras claves extraídas del propio documento. Estas aplicaciones constituyen los principales procesos de la descripción de contenido, es decir, la clasificación y la indización.

Sobre la base de lo anterior, Dahlberg aporta la siguiente definición:

La Organización del Conocimiento es la ciencia de estructurar y ordenar sistemáticamente las unidades de conocimiento (conceptos) de acuerdo

con sus elementos de conocimiento inherentes (características) y la aplicación de conceptos y clases de conceptos ordenados por esta vía para la asignación de contenidos de valor de los referentes (objetos/sujetos) de todas clases (2006b).

En otro artículo, Dalhberg (1993) establece el alcance de la organización del conocimiento mediante un esquema de conceptos que se reproduce en el Apéndice A de este capítulo. A pesar del tiempo transcurrido y de la necesidad de incluir algunas novedades como los mapas conceptuales y las ontologías, el esquema de Dahlberg ayuda a comprender de qué se ocupa esta disciplina.

Por otro lado, la misma Dahlberg (2006a) estima que la organización del conocimiento puede considerarse una nueva ciencia y la relaciona con el sistema de clasificación *Information Coding Classification*, elaborado por ella misma en 1982. De acuerdo con esto, la ubica en la clase 81 Ciencia de la Ciencia y, específicamente, dentro de la 814 Sistemología de la Ciencia de la Ciencia. Esto nos recuerda al *Broad System of Ordering* publicado hace más de tres décadas por la Federación Internacional de Documentación y la Unesco, de modo que podría incluirse en la clase 116 Ciencia de la Ciencia y del Conocimiento (Coates *et al.*, 1981).

Otros autores también han propuesto definiciones de la organización del conocimiento. El profesor Birger Hjørland de la Aalborg Universitet (Dinamarca) define la disciplina en dos sentidos:

En un sentido estricto, la Organización del Conocimiento (OC) trata sobre actividades tales como la descripción documental, la indización y la clasificación llevada a cabo en bibliotecas, bases de datos bibliográficas, archivos y otros tipos de instituciones de la memoria por bibliotecarios, archivistas, especialistas de información, especialistas en materias, así como por algoritmos de computación y personas comunes. La Organización del Conocimiento como campo de estudio se ocupa de la naturaleza y calidad de tales procesos organizadores del conocimiento, así como de los sistemas organizadores del conocimiento usados para organizar documentos, representaciones documentales, obras y conceptos.

(...)

En un sentido amplio, la OC trata sobre la división social del trabajo

mental, es decir, la organización de universidades y otras instituciones de investigación y educación superior, la estructura de las disciplinas y profesiones, la organización social de los medios de producción y la diseminación del conocimiento (Hjørland, 2008a: 86).

La definición de Hjørland, en sentido estricto, abarca tanto los SOC como los procesos de organización del conocimiento (POC) y resalta varios aspectos: a) el ámbito en el que se llevan a cabo (bibliotecas, bases de datos, archivos, etc., a los que en conjunto denomina *instituciones de la memoria*); b) quiénes los llevan a cabo (bibliotecarios, archivistas, especialistas de la información) y c) el aporte de los usuarios y de la tecnología informática.

En cuanto a la definición en un sentido amplio, Hjørland no se limita -como Bliss- al conocimiento desarrollado en las instituciones académicas, sino que se extiende también hacia otros organismos productores de conocimiento que actúan en la sociedad, como las asociaciones profesionales, las empresas, agencias de gobierno, organismos no gubernamentales, etc. (La organización del conocimiento en sentido amplio escapa al alcance de esta obra).

Broughton *et al.* -entre los que se encuentra el propio Hjørland- también distinguen entre el sentido estricto y el sentido amplio de la organización del conocimiento. Respecto al sentido estricto aportan la siguiente definición:

La Organización del Conocimiento (OC), en sentido estricto, trata sobre los sistemas organizadores del conocimiento (SOC), tales como los registros bibliográficos, los sistemas de clasificación (por ejemplo SCDD, LCC y CDU), los tesauros, las redes semánticas y sobre los procesos organizadores del conocimiento tales como la clasificación, la descripción documental, la “catalogación descriptiva”, la indización y el análisis por materia. La OC se lleva a cabo en instituciones de la memoria, tales como bibliotecas, archivos, museos y bases de datos en línea y en Internet, pero también fuera de tales instituciones, por ejemplo en los índices analíticos de los libros y los llamados sistemas de gestión de información personales [...] La asignación de términos, textos y notaciones a los registros (o a los objetos mismos como en los recursos digitales), está relacionada con el contenido intelectual o semántico de los recursos y es independiente del formato en que la información es presentada. La OC también se inte-

resa por el diseño de SOC y los principios y metodologías para construir herramientas semánticas (Broughton *et al.*, 2005: 133-134).

Al igual que la anterior, esta definición abarca los SOC y los POC, pero extiende el ámbito de la organización del conocimiento, en sentido estricto, más allá de las instituciones de la BDCI, para incluir tanto la industria editorial que produce los índices analíticos de libros y revistas como los sistemas de gestión personales. Señala que los POC se aplican a los registros bibliográficos, pero también a los registros metadatos que se incrustan o vinculan a los recursos digitales disponibles en la Web. Tanto Hjørland como Broughton *et al* incluyen dentro de los POC a los procesos de descripción de contenido (clasificación, indización, etc.), pero no son muy explícitos en cuanto al diseño y desarrollo de SOC, que, como ya se mencionó, también es un conjunto de procesos de organización del conocimiento.

El profesor Claudio Gnoli, de la Università degli studi di Pavia (Italia), propone un modelo tentativo para estudiar la organización del conocimiento, que se basa en una estructura formada por cuatro *capas* (Gnoli, 2011). La primera capa se refiere a las teorías y abordajes del conocimiento que provienen de otras disciplinas como la filosofía de la ciencia, la psicología, la sociología, etc. La segunda capa abarca los SOC. La tercera capa, que Gnoli llama de *representaciones*, incluye las normas y formatos de naturaleza técnica como MARC, RDF, SKOS u OWL, que permiten representar e intercambiar datos en entornos digitales y que constituyen la gran novedad del tramo final del siglo veinte y del presente. Por último, la cuarta capa se refiere a la aplicación de las teorías, los SOC y las representaciones a los directorios, catálogos, bibliografías, etc. de archivos, bibliotecas, museos u otras instituciones similares; en otras palabras, la cuarta capa se ocupa de los POC.

Barité *et al.*, en su *Diccionario de organización del conocimiento*, aportan la siguiente definición:

Área del conocimiento de formación reciente, que estudia las leyes, los principios y los procedimientos por los cuales se estructura el conocimiento especializado en cualquier disciplina, con la finalidad de representar temáticamente y recuperar la información contenida en documentos de cualquier índole, por medios eficientes que den respuesta rápida a las

necesidades de los usuarios. La Organización del Conocimiento se nutre de los aportes recibidos de la Informática, la Lingüística, la Terminología y las Ciencias de la Información. García Marco (1995) señala que es “una disciplina científica (esto es, sujeta a método científico) y social, de carácter aplicado” pues tiene como objetivo “optimizar la circulación del conocimiento en las sociedades”. El objeto de estudio de la Organización del Conocimiento es el conocimiento socializado o registrado, y en lo que hace a la Bibliotecología y Documentación, da cuenta del desarrollo teórico-práctico para la construcción, la gestión, el uso y la evaluación de clasificaciones, taxonomías, nomenclaturas, ontologías temáticas y lenguajes documentales. Asimismo, ampara el conjunto de conocimientos vinculados al análisis de información en general, considerando aspectos semánticos, cognitivos e informáticos (Barité *et al.*, 2013).

Para estos autores, la organización del conocimiento abarca la teoría (leyes, principios, conjunto de conocimientos) y se ocupa tanto de los SOC -a los que denominan *lenguajes documentales*- como de los POC (construcción, gestión, uso y evaluación de los SOC). Además, reconocen el carácter interdisciplinar de la organización del conocimiento (aspectos semánticos, cognitivos e informáticos), pero destacan que ocupa un lugar en la BDCI. Cumple también con una función (representar y recuperar información, dar respuesta al usuario) y adhieren a la concepción de García Marco en cuanto se trata de una ciencia social de carácter aplicado.

Por su parte, el profesor Miguel Ángel Esteban Navarro de la Universidad de Zaragoza (España) propone otra definición:

La Organización del Conocimiento es la disciplina específica que se dedica dentro de la Ciencia de la Información Documental al estudio de los fundamentos teóricos del tratamiento y la recuperación de la información y a la construcción, mantenimiento, uso y evaluación de los instrumentos lógico-lingüísticos más adecuados para controlar los procesos de recuperación, clasificación, ordenación y almacenamiento del contenido informativo de los documentos con el fin de permitir su recuperación y comunicación. Se ocupa, por tanto, de los principios y herramientas

puestos en acción para la gestión del conocimiento humano desde una triple perspectiva: su representación, su organización y su comunicación documental (Esteban Navarro, 1996: 97-98).

En esta definición también se hace referencia a la teoría, a los SOC (instrumentos lógico-lingüísticos) y a los POC (construcción de los SOC, recuperación, clasificación, ordenación, almacenamiento), y se considera que la organización del conocimiento es una disciplina dentro de lo que este autor denomina ciencia de la información documental.

Más allá de estas definiciones, en 1995 el mismo Esteban Navarro consideraba, junto al profesor Francisco Javier García Marco de la Universidad de Zaragoza, que la organización del conocimiento es una disciplina en formación. En nuestra opinión, si bien se ha avanzado en distintos aspectos, esta afirmación sigue siendo válida:

La Organización del Conocimiento se encuentra en la actualidad ante dos retos comunes a todas las disciplinas científicas durante su nacimiento y consolidación. Por una parte, debe definir sus fundamentos gnoseológicos, establecer sus principios teóricos y metodológicos y fijar sus límites dentro del mapa de las ciencias para diseñar el tipo de relaciones y de intercambio de conocimientos y de métodos que mantendrá con las disciplinas concurrentes en su formación o próximas. Y, por otra parte, debe acotar su campo de acción en la praxis, construir una teoría adecuada para analizar la realidad sobre la que actúa y guiar el ejercicio de su práctica en ella, y, como complemento, desarrollar los instrumentos técnicos necesarios para ejercer adecuadamente esa dimensión práctica. En definitiva, se encuentra generando una tecnología (Esteban Navarro & García Marco, 1995: 149-150).

La organización del conocimiento y las corrientes de la BDCI

La organización del conocimiento se nutre de otras disciplinas como la epistemología, la psicología cognitiva, la sociología del conocimiento y, particularmente, de la conceptología, la lingüística, la terminología, la informática y la teoría de la comunicación (García Marco, 1995; Guimarães, 2001a).

Sin embargo, la BDCI es considerada como “la disciplina central de la Or-

ganización del Conocimiento en este sentido estricto” (Hjørland, 2008a: 86) y, por tal motivo, se ha propuesto a la organización del conocimiento como marco teórico referencial para aquellas actividades de la BDCI relacionadas con el tratamiento temático del conocimiento explícito o registrado (Barité, 2001).

En varias publicaciones, Guimarães ha procurado encuadrar las distintas corrientes de la BDCI dentro de la organización del conocimiento (Guimarães, 2001b, 2008, 2009; Guimarães *et al.*, 2012). Así, reconoce tres corrientes:

- Corriente bibliotecológica de origen estadounidense (catalogación de materia).
- Corriente documental de origen francés (análisis de contenido, análisis formal de contenido, descripción de contenido).
- Corriente informacional de origen británico (indización).

El autor destaca que estas corrientes comparten una historia común y estudian fenómenos semejantes a pesar de las distintas denominaciones (Guimarães, 2009) y concluye que “la organización del conocimiento actúa como espacio de agregación entre las demás corrientes, contribuyendo significativamente para el diálogo entre las mismas” (Guimarães *et al.*, 2012).

Ámbito de la organización del conocimiento

Finalmente, puede afirmarse que la organización del conocimiento ya no es patrimonio exclusivo de las bibliotecas, centros de documentación y otras instituciones de la BDCI, sino que se ha extendido a otros organismos que gestionan conocimiento —como las empresas, agencias de gobierno u otros organismos—, y a la World Wide Web. Como afirma María Inés Cordeiro, ex directora de la Biblioteca Nacional de Portugal y asesora del Consorcio CDU:

La Organización del Conocimiento ha sido siempre reconocida como una de las funciones primarias de las bibliotecas, que han sido uno de los principales proveedores de este servicio. Los profundos cambios introducidos por Internet han incrementado el interés acerca del conocimiento y de la Organización del Conocimiento, diversificando las perspectivas acerca de la gestión de conocimiento. El campo se ha vuelto más complejo y menos claro como campo profesional. De repente, la Organización del Conocimiento se encuentra a sí misma en la múltiple encrucijada

del paisaje sobrecargado y distribuido de la información generado por la WWW (Cordeiro, 2003: 66).

Organización del conocimiento vs. organización de la información

Algunos especialistas limitan la organización del conocimiento al diseño y desarrollo de SOC, pero consideran que la descripción de contenido es incumbencia de la ciencia de la información (Bräscher & Café, 2008; Café *et al.*, 2012).

Otros, por el contrario, consideran que esta disciplina se ocupa solamente de la descripción de contenido. Por ejemplo, el profesor emérito James D. Anderson, de la Rutgers University (Estados Unidos), propone una definición que toma en cuenta la indización, la clasificación, etc., pero no hace referencia al diseño y desarrollo de los SOC necesarios para llevar a cabo esos procesos:

La Organización del Conocimiento en Bibliotecología y Ciencia de la Información consiste en diversos aspectos: identificación de los mensajes (con frecuencia llamados *obras*); identificación de los textos en los cuales se representan los mensajes (*obras*); descripción de los documentos en los cuales se presentan los textos; y descripción del contenido, características y significado de los mensajes. Las identificaciones y descripciones resultantes se organizan en índices, catálogos, bases de datos, bibliotecas digitales y otros sistemas de recuperación de información para el acceso de los interesados. La creación y organización de estas identificaciones y descripciones se llaman indización, condensación, catalogación, clasificación bibliográfica, bibliografía, gestión de registros y gestión de conocimiento, en varios contextos (Anderson, 2003: 471-472).

Otros autores rechazan terminantemente el concepto ‘organización del conocimiento’ y prefieren ‘organización de la información’, como la profesora emérita de la University of Pittsburg (Estados Unidos) Arlene G. Taylor y el profesor emérito de la University of Sheffield (Gran Bretaña) Tom D. Wilson:

El conocimiento existe en la mente de un individuo que ha estudiado una

materia, la ha comprendido y quizás ha contribuido con algo a través de la investigación u otros medios [...] la información es la comunicación o recepción de conocimiento [...] Así, elegí usar el término *información* en lugar de *conocimiento* como expresión de lo que creo que nosotros organizamos cuando organizamos en beneficio de otras personas (Taylor, 2004: 3).

El conocimiento involucra los procesos mentales de comprensión, entendimiento y aprendizaje que ocurren en la mente y solo en la mente [...] Siempre que deseamos expresar lo que conocemos, solo podemos hacerlo por mensajes de uno u otro tipo: orales, escritos, gráficos, gestuales o incluso a través del lenguaje corporal. Tales mensajes no portan conocimiento, ellos constituyen información que una mente cognoscente puede asimilar, entender, comprender e incorporar a sus propias estructuras de conocimiento [...] La consecuencia de este análisis es que todo lo externo a la mente que pueda ser manipulado en cualquier forma, puede definirse como ‘datos’, si consiste en simples hechos, o como ‘información’, si los datos están incorporados en un contexto de relevancia para el receptor [...] De modo que los datos y la información pueden ser gestionados, los recursos de información pueden ser gestionados, pero el conocimiento (es decir, lo que conocemos) nunca puede ser gestionado, excepto por el individuo que conoce y, aun así, solo imperfectamente (Wilson, 2002).

Representación y organización del conocimiento

Algunos autores, en particular del ámbito iberoamericano, prefieren el término *representación y organización del conocimiento* (Esteban Navarro, 1996; Esteban Navarro & García Marco, 1995) o bien al revés, *organización y representación del conocimiento* (Fujita, 2008; Guimarães *et al*, 2012; Peña Vera, 2010, 2011). La revista *Scire: representación y Organización del Conocimiento*,¹ publicada por la Facultad de Filosofía y Letras de la Universidad de Zaragoza, España, hace referencia al término en cuestión tanto en el subtítulo como en su política editorial.

Según la profesora venezolana Tania Peña Vera (2010, 2011) los procesos de representación incluyen la descripción, la catalogación, la indización

¹ Scire: <http://www.ibersid.eu/ojs/index.php/scire/index>

y la condensación, mientras que para Esteban Navarro (1996) comprenden la indización y la condensación. Ambos autores coinciden en que los procesos de organización son la clasificación y la ordenación.

Estos procesos de representación se limitan a los POC de la descripción de contenido, pero el diseño y desarrollo de SOC también incluye otros procesos de representación. Por eso, adherimos a la posición de Dahlberg (2006a), para quien la organización del conocimiento comprende la representación como una forma necesaria para comunicar el conocimiento individual que es subjetivo, compararlo con las representaciones de otras personas y contrastarlo con la realidad para verificar su verdad. La representación del conocimiento -afirma Dahlberg- se ocupa tanto de la estructura lógica de los conceptos (conceptología) como de la designación de estos (terminología) y distingue cuatro niveles de representación, de lo más simple a lo más complejo:

- Elementos de conocimiento: son las características de los conceptos a partir de las cuales se pueden predicar las propiedades o hacer declaraciones sobre los objetos. Las características son abstracciones de las propiedades de los objetos del mundo real o del ideal, que se expresan mediante enunciados.
- Unidades de conocimiento o conceptos: son representaciones mentales de los objetos del mundo real o del ideal y constituyen la síntesis de las características. Se expresan mediante designaciones verbales o simbólicas como los términos preferentes de un tesoro o las notaciones de un sistema de clasificación.
- Unidades de conocimiento amplias: son combinaciones de conceptos, como pueden ser las declaraciones, las definiciones y los textos.
- Sistemas de conocimiento: son estructuras planificadas y cohesivas que ordenan los conceptos de un determinado dominio, por ejemplo un sistema de clasificación o un tesoro.

Entonces, hay representación del conocimiento en la indización y la condensación, que son procesos de la descripción de contenido, pero también hay representación en un tesoro, en un sistema de clasificación o en otros SOC.

La misma Dahlberg (2006a) considera que la representación del conocimiento es parte de la organización del conocimiento y por lo tanto basta con este último término, que en definitiva es el que se ha generalizado. De nuevo

Esteban Navarro & García Marco (1995) ofrecen una explicación:

Se trata, por tanto, de una ciencia tridimensional, ya que se ocupa de los principios, métodos e instrumentos puestos en acción para la gestión del conocimiento humano desde una triple perspectiva: su representación, su organización y su comunicación documental. No obstante, la Ciencia de la Representación, Organización y Comunicación del Conocimiento se denomina de modo más simple y breve Organización del Conocimiento, debido a que la organización es el elemento mediador entre los otros dos actos, ya que, por un lado, la representación se efectúa con la finalidad de permitir una eficaz organización, y, por otro, la comunicación exige una correcta recuperación, cuyo éxito depende de la calidad de la organización (Esteban Navarro & García Marco, 1995: 149).

Organización del conocimiento y gestión del conocimiento

El término *organización del conocimiento* no debe confundirse con otro término de amplia difusión en años recientes que es *gestión del conocimiento*. Este último se refiere a una nueva forma de gestión para las empresas u otros organismos, basada en el reconocimiento de que su principal activo es el capital intelectual, es decir, el conocimiento que poseen las personas que allí se desempeñan. La gestión del conocimiento es:

Un proceso sistémico y específico de cada organismo para adquirir, organizar y comunicar el conocimiento de los empleados, de modo que otros empleados puedan hacer uso de ese conocimiento para ser más efectivos y productivos en su trabajo (Gottschalk, 2007: 25).

En este marco y a partir del trabajo pionero de Nonaka y Takeuchi, se ha distinguido entre conocimiento tácito y conocimiento explícito (Choo, 2003). Como ya se mencionó, el conocimiento tácito es aquel que existe en la mente de las personas y se manifiesta a través de sus acciones. El conocimiento explícito es el que ha sido formalizado y documentado en manuales de procedimientos, especificaciones técnicas, instructivos, normas, etc.

Un sistema de gestión de conocimiento debe procurar que la mayor cantidad posible de conocimiento tácito se convierta en conocimiento explícito,

porque es este último el que puede ser gestionado. En este sentido, Choo (2003) señala que uno de los proyectos más importantes de la gestión del conocimiento es la creación de repositorios institucionales que tienen como objetivo capturar el conocimiento registrado y organizarlo de modo que pueda ser recuperado fácil y rápidamente.

En consecuencia, dentro de un sistema de gestión del conocimiento, la organización del conocimiento es útil para: a) la descripción del contenido de aquellos documentos que formarán parte del repositorio institucional y b) el diseño y desarrollo de SOC específicos, particularmente tesauros, taxonomías y ontologías que serán aplicados al sitio Web y la intranet del organismo.

De este modo, podemos decir que en el ámbito de un organismo público o privado, la organización del conocimiento puede ser de utilidad para la gestión del conocimiento, de una manera similar -aunque no idéntica- a lo que ocurre en una biblioteca.

Autoevaluación

En el ámbito de la bibliotecología, la documentación y la ciencia de la información, ¿quién fue el primero en utilizar el término *organización del conocimiento*? ¿Quiénes retomaron el término en la segunda mitad del siglo veinte?

¿Cuál es el objeto de estudio de la organización del conocimiento?

¿De qué se ocupa la organización del conocimiento?

¿Cuál es su función?

¿Quién la lleva a cabo?

¿En qué ámbitos se lleva a cabo?

La organización del conocimiento, ¿es una ciencia, una disciplina, una especialidad o un campo de conocimiento?

Con las respuestas de las preguntas b a g redacte su propia definición de organización del conocimiento.

¿Cuál es la diferencia entre organización del conocimiento y gestión del conocimiento?

¿De qué manera puede contribuir la organización del conocimiento a la gestión del conocimiento?

Apéndice A. Alcance de la organización del conocimiento²

0 Divisiones de forma

- 01 Bibliografías
- 02 Revisiones bibliográficas
- 03 Diccionarios, terminologías
- 04 Sistemas de clasificación y tesauros
- 05 Publicaciones periódicas y series
- 06 Actas de congresos
- 07 Libros de texto
- 08 Otras monografías
- 09 Normas, directrices

1 Fundamentos teóricos y problemas generales

- 11 Orden y organización del conocimiento
- 12 Conceptología y organización del conocimiento
- 14 Teoría de sistemas y organización del conocimiento
- 15 Psicología y organización del conocimiento
- 16 Ciencia y organización del conocimiento
- 17 Problemas en organización del conocimiento
- 18 Investigación de la organización del conocimiento
- 19 Historia de la organización del conocimiento

2 Sistemas de clasificación y tesauros: estructura y construcción

- 21 Cuestiones generales de sistemas de clasificación y tesauros
- 22 Estructura y elementos de sistemas de clasificación y tesauros
- 23 Construcción de sistemas de clasificación y tesauros
- 24 Relaciones
- 25 Taxonomía numérica
- 26 Notación. Códigos
- 27 Mantenimiento, actualización y almacenamiento de sistemas de clasificación y tesauros
- 28 Compatibilidad y concordancia entre sistemas de clasificación y tesauros

² Traducido de Dahlberg, 1993.

29 Evaluación de sistemas de clasificación y tesauros

3 Clasificación e indización: metodología

31 Teoría de clasificación e indización

32 Análisis por materia

33 Técnicas de clasificación e indización

34 Clasificación e indización automáticas

35 Ordenamiento manual y automatizado

36 Codificación

37 Reclasificación

38 Generación de índices

39 Evaluación de la clasificación y la indización

4 Sistemas de clasificación y tesauros universales

41 Sistemas de clasificación universales en general

42 Clasificación decimal universal

43 Sistema de clasificación decimal de Dewey

44 Clasificación y epígrafes de Library of Congress

45 Bliss bibliographic classification

46 Colon classification

47 Library bibliographical classification

48 Otros sistemas de clasificación y tesauros universales

49 (Vacía)

5 Taxonomías especializadas en

51 Forma y estructura

52 Energía y materia

53 Cosmos & geografía

54 Biología

55 Humanidades

56 Ciencias sociales

57 Economía y tecnología

58 Ciencia e información

59 Cultura

6 Sistemas de clasificación y tesauros especializados en

- 61 Forma y estructura
- 62 Energía y materia
- 63 Cosmos & geografía
- 64 Biología
- 65 Humanidades
- 66 Ciencias sociales
- 67 Economía y tecnología
- 68 Ciencia e información
- 69 Cultura

7 Representación del conocimiento por lengua natural y terminología

- 71 Problemas generales de la lengua natural respecto a la organización del conocimiento
- 72 Semántica
- 73 Procesamiento automático de la lengua
- 74 Problemas gramaticales
- 75 Sistemas y tecnología de recuperación en línea
- 76 Problemas de léxicos/diccionarios
- 77 Problemas de terminología
- 78 Terminología orientada por materia
- 79 Problemas de sistemas multilingües y traducción

8 Clasificación e indización aplicada

- 81 Problemas generales, catálogos, directrices, reglas, índices
- 82 Clasificación e indización de datos
- 83 Clasificación e indización de títulos
- 84 Clasificación e indización de información primaria (excepto 85)
- 85 Clasificación e indización del texto de los libros
- 86 Clasificación e indización de información secundaria
- 87 Clasificación e indización de materiales distintos al libro
- 88 Clasificación e indización en campos temáticos
- 89 Clasificación e indización en ciertas lenguas

9 Ambiente de la organización del conocimiento

- 91 Problemas profesionales y organizacionales
- 92 Personas e instituciones en la organización del conocimiento
- 93 Organizaciones de clasificación e indización en el nivel nacional e internacional
- 94 (Vacía)
- 95 Educación y entrenamiento en organización del conocimiento
- 96 Cuestiones legales en organización del conocimiento
- 97 Aspectos económicos en organización del conocimiento
- 98 Estudios de usuarios y organización del conocimiento
- 99 Normalización en organización del conocimiento

Sistemas y procesos de organización del conocimiento

Como se expuso en el capítulo1, la organización del conocimiento se ocupa de sistemas y procesos, que trataremos de explicar a continuación, así como el uso de la lengua natural en los procesos de organización y recuperación del conocimiento registrado.

Sistemas de organización del conocimiento

A lo largo de los años se han acuñado diferentes y numerosos términos para designar el mismo o casi el mismo concepto: lenguajes de indización, lenguajes documentales, lenguajes documentarios, lenguajes de recuperación de información, lenguajes de materias, lenguajes terminológicos, vocabularios controlados, vocabularios estructurados, vocabularios controlados y estructurados (IFLA, 2011: 84). A estos términos se han sumado otros más recientes: sistema de organización del conocimiento (Hjerppe, 1990) y sistema organizador del conocimiento (Hjørland, 2008a). También se han propuesto sistema de representación y organización del conocimiento (Esteban Navarro, 1996), instrumento de organización del conocimiento (Fujita, 2008) y herramienta de organización del conocimiento (Neelamegan & Lalitha, 2013).

Esta diversidad terminológica es producto de distintas corrientes y concepciones de los especialistas en el tema. Aquí elegimos el término *sistema de organización del conocimiento* porque se relaciona directamente con la disciplina a la que nos referimos y porque ha logrado mayor difusión en la última década. Este término comenzó a utilizarse en la década de 1990 (Hjerppe, 1990, 1996), pero parece haber obtenido más repercusión a partir del primer congreso de la ACM Digital Libraries, que se llevó a cabo en Pittsburgh, Estados Unidos, en 1998 (Hodge, 2000). Desde entonces, se han

hecho intentos por definirlo, aunque todavía no se ha llegado a un acuerdo y de hecho existen distintas posiciones entre los expertos.

Una primera posición, tal vez la más difundida, es la de aquellos que han puesto el énfasis en establecer una tipología de los SOC y se conforman con una definición por extensión, que abarque los distintos tipos. En el año 2000, la especialista estadounidense Gail Hodge preparó un informe para la Digital Library Federation, avalado por el Council on Library and Information Resources de su país, en el que aporta la siguiente definición de SOC:

El término *sistemas de organización del conocimiento* intenta abarcar todos los tipos de esquemas para organizar la información y promover la gestión de conocimiento. Los sistemas de organización del conocimiento incluyen los esquemas de clasificación y categorización que organizan los materiales en un nivel general, los encabezamientos de materia que proporcionan acceso más detallado y los archivos de autoridades que controlan las versiones variantes de la información clave como los nombres geográficos y personales. Los sistemas de organización del conocimiento también incluyen vocabularios altamente estructurados, como los tesauros, y los esquemas menos tradicionales, como las redes semánticas y las ontologías. Debido a que los sistemas de organización del conocimiento son mecanismos para organizar información, están en el corazón de cada biblioteca, museo y archivo (Hodge, 2000: 1).

Como se puede observar, la definición anterior no enuncia las características de los SOC sino que solamente enumera los distintos tipos que abarca este término. En su informe, Hodge propone una tipología de los SOC que incluye las siguientes categorías:

- Listas de términos: listas de autoridades, glosarios, diccionarios, listas de nombres geográficos.
- Clasificaciones y categorías: encabezamientos de materia, sistemas de clasificación.
- Listas de relaciones: tesauros, redes semánticas, ontologías.

Otros autores europeos, basados en la definición de Hodge, también consideran que el término *sistema de organización del conocimiento* comprende

tesauros, listas de términos, ontologías, sistemas de clasificación, taxonomías, listas de nombres geográficos, diccionarios, bases de datos léxicas, mapas conceptuales y espaciales, mapas de ruta semánticos, etc. (Hill & Koch, 2001; Tudhope & Koch, 2004).

En 2006, los profesores Douglas Tudhope de la University of Glamorgan (Gran Bretaña), Traugott Koch de la Lunds Universitet (Suecia) y la experta Rachel Heery de UKOLN (Gran Bretaña) presentaron un informe al Joint Information System Committee de Gran Bretaña, en el que clasifican los SOC por su estructura y su propósito (Tudhope *et al.*, 2006).

De acuerdo con su estructura, estos autores clasifican los SOC como sigue:

- Listas de términos no estructuradas: listas de selección (*pick lists*), listas de autoridades, glosarios, diccionarios, anillos de sinónimos.
- Taxonomías: sistemas taxonómicos científicos, taxonomías Web, taxonomías corporativas.
- Encabezamientos de materia: listas de epígrafes generales como LCSH o especializadas como MeSH.
- SOC basados en relaciones: tesauros, sistemas de clasificación, bases de datos léxicas (por ejemplo WordNet) y ontologías.

Los mismos autores agrupan los SOC de acuerdo con su función de la siguiente manera:

- Con fines de recuperación: SOC destinados a la clasificación e indización en catálogos de bibliotecas, bibliografías, bibliotecas digitales, etc.
- Con fines lingüísticos: los SOC utilizados en la traducción asistida por computadora, la ingeniería de lenguaje, la minería de textos y las bases de datos léxicas como WordNet.
- Con fines de inteligencia artificial: las ontologías.
- Con fines de e-aprendizaje (*e-learning*): los metadatos para objetos de aprendizaje y sus repositorios, muchas veces ligados a las bibliotecas digitales.
- Con fines científicos: las taxonomías y nomenclaturas científicas, entre otros.

La profesora Marcia Lei Zeng de la Kent State University (Estados Unidos), basándose también en el trabajo de Hodge y el de Tudhope *et al.*, propone cuatro categorías para clasificar los SOC por su estructura (Zeng, 2008):

- Listas de términos: incluye las listas de selección, diccionarios, glosarios y anillos de sinónimos.
- Modelos similares a metadatos: listas de autoridades, directorios, listas de nombres geográficos.
- Clasificación y categorización: encabezamientos de materia, esquemas de categorización, taxonomías, sistemas de clasificación.
- Modelos de relaciones: tesauros, redes semánticas, ontologías.

Tabla 1 - Características de los SOC según sus funciones

SOC	Control de ambigüedad	Control de sinónimos	Relaciones jerárquicas	Relaciones asociativas	Propiedades
Glosarios, diccionarios, listas de términos	xxx	-	-	-	-
Anillos de sinónimos	-	xxxx	-	-	-
Listas de autoridades, directorios, listas de nombres geográficos	xxx	xxx	x	-	-
Sistemas de clasificación, esquemas de categorización, taxonomías, listas de epígrafes	xx	xx	xxxx	-	-
Tesauros	xxxx	xxxx	xxx	xxxx	-
Ontologías	xx	xx	xxx	xxxxx	xxxx

Zeng reconoce cinco funciones de los SOC que se muestran en la Tabla 1. Como se puede observar, no todos los SOC cumplen con todas las funciones o no lo hacen en el mismo grado.

Por su parte, los profesores Renato Rocha Souza y Mauricio Barcellos Almeida, de la Universidade Federal de Minas Gerais (Brasil), junto al ya mencionado Douglas Tudhope, han analizado en sucesivos artículos distintas propuestas de otros autores, a partir de lo cual presentan una tipología de los SOC con un criterio amplio, que no es mutuamente excluyente, de modo que algunos SOC han sido clasificados en más de una categoría (Souza *et al.*, 2010, 2012). Según el tipo de estructura, los SOC se dividen de la siguiente manera:

- Texto no estructurado: resúmenes, surrogados.
- Listas de términos y/o conceptos: folksonomías, diccionarios, glosarios, listas de autoridades, listas de etiquetas, anillos de sinónimos, líneas de concordancia, vocabularios controlados.
- Estructuras de conceptos y relaciones: vocabularios controlados, sistemas de clasificación, esquemas de categorización, encabezamientos de materia, índices para la recuperación de información, diccionarios de datos, taxonomías, tesauros, redes semánticas, ontologías.
- Estructuras de conceptos, relaciones y distribución: mapas mentales, mapas argumentales, mapas conceptuales, *rich pictures* (representaciones de conceptos mediante imágenes), modelos entidad-relación, modelos referenciales, modelos de datos.

Las tipologías anteriores son propuestas y, por lo tanto, están sujetas a discusión, ya que de hecho no coinciden totalmente entre ellas, ni se ajustan a una definición precisa que exprese sin ambigüedad las características esenciales de un SOC. El tema requiere un análisis más profundo que excede el alcance de esta obra.

En una segunda posición encontramos a Birger Hjørland, mencionado anteriormente, quien utiliza en sus textos dos términos similares, pero no idénticos, para referirse a los SOC: a) *knowledge organization systems*, que puede traducirse como sistemas de organización del conocimiento (Hjørland, 2008b) y b) *knowledge organizing systems* o sistemas organizadores del conocimiento (Hjørland, 2008a). Para Hjørland se trata de un “término general que se refiere, entre otras cosas a las herramientas que presentan una interpretación organizada de las estructuras del conocimiento” (Hjørland, 2008b). También analiza los SOC en un sentido amplio y en un sentido estricto (Figura 1).

En el sentido amplio de los SOC, Hjørland incluye las bibliografías, las teorías y sistemas conceptuales, las culturas y subculturas, las disciplinas y la división social del trabajo, las enciclopedias, los medios de comunicación, las lenguas y los sistemas simbólicos, los modelos de actores, sistemas y procesos en diferentes dominios, etc. En un sentido estricto, es decir, en el área de la BDCI, Hjørland agrupa ciertos SOC a los que denomina *herramientas semánticas* y postula la siguiente definición:

Una herramienta semántica es una herramienta que proporciona información semántica, esto es información acerca del significado de las palabras y otros símbolos, así como las relaciones entre símbolos y conceptos (relaciones semánticas) (Hjørland, 2007b).

Figura 1. Sistemas de organización del conocimiento (SOC), según Hjørland (2006-2008b).



Entre las herramientas semánticas, Hjørland incluye los lenguajes de indización, que define como “lenguas usadas para la clasificación o la indización de las materias de los documentos” (Hjørland, 2006) y los divide en dos grandes grupos: a) los sistemas de clasificación y b) los lenguajes de indización verbales, ya sean libres o controlados (tesauros, listas de epígrafes, etc.), que se distinguen según su forma de coordinación. Dentro de las herramientas semánticas se incluyen también los mapas conceptuales, los hipertextos, los anillos de sinónimos, tipologías, folksonomías y aquellos otros incluidos en la tipología de Hodge.

Según Hjørland, las herramientas semánticas tienen semejanzas y diferencias. Por un lado, comparten dos características: a) de una u otra forma despliegan un conjunto de símbolos y conceptos, y b) proporcionan información acerca de las relaciones semánticas entre los conceptos desplegados. Así, los conceptos son las unidades de conocimiento que se organizan, y la manera en que se organizan es de acuerdo con su significado, es decir, de

acuerdo con las relaciones semánticas entre los conceptos. Por otro lado, las herramientas semánticas se diferencian entre sí por la clase de símbolos y los conceptos seleccionados y por la información proporcionada acerca de cada concepto, en particular sus relaciones semánticas.

Broughton *et al.* coinciden en su preferencia por el término *sistemas organizadores del conocimiento* y en el enfoque sobre el sentido amplio y el sentido estricto de los SOC. Definen el término en sentido estricto como “tipos de herramientas semánticas que proporcionan la selección de conceptos e información acerca de sus relaciones semánticas” (Broughton *et al.*, 2005: 142) e incluyen los siguientes:

- Sistemas de clasificación como SCDD, CDU, LCC
- Sistemas de clasificación facetados como la segunda edición de BC
- Listas de epígrafes como LCSH
- Sistemas basados en la búsqueda por texto libre
- Tesoros
- Mapas bibliométricos
- Algoritmos de los motores de búsqueda
- Sistemas de archivo (basados en el principio de procedencia)
- Ontologías
- Redes semánticas
- Mapas de tópicos

De modo que para los autores que proponen esta segunda posición, *sistema de organización del conocimiento* es un concepto amplio que incluye —entre otros, pero no exclusivamente— las herramientas semánticas, y son estas últimas las que coinciden, al menos en parte, con las tipologías de los SOC propuestas por los autores de la primera posición.

Otros autores iberoamericanos consideran que *sistema de organización del conocimiento* es lo mismo que lenguaje documental o lenguaje documentario:

Los SOC también llamados “lenguajes documentales”, “sistemas organizadores del conocimiento” o “sistemas de organización y representación del conocimiento” [...] (Barité *et al.*, 2013).

Los lenguajes documentarios, caracterizados como sistemas de organiza-

ción del conocimiento y correspondientes a las listas de encabezamientos de materia y a los tesauros, tienen como primera función representar el contenido de los documentos contenidos en un sistema de recuperación de información función por el contenido y, como segunda función, mediar la recuperación de información por medio de la representación de las preguntas formuladas por los usuarios, función por el uso (Boccatto, 2009: 120).

También nosotros hemos afirmado anteriormente que a los SOC “se los ha conocido como lenguajes documentales” (Martínez Tamayo & Valdez, 2009: 28), pero debe tenerse en cuenta que los SOC van más allá de los tradicionales sistemas de clasificación, listas de epígrafes y tesauros y que exceden el campo de la BDCI.

Finalmente, llama la atención que ninguna de las definiciones y tipologías anteriores haga referencia a que un SOC es, precisamente, un sistema. Por ende no declaran, de manera explícita y completa, cuáles son los elementos componentes de este sistema, cómo se interrelacionan y estructuran, ni cuál es su propósito. En ese sentido, recordamos que Dahlberg (2006a) los llama *sistemas de conceptos* y coincidimos con la recomendación de Esteban Navarro (1996) respecto a que la definición del término debe basarse en el análisis de las palabras que lo forman.

Así, en esta obra definimos el término *sistema de organización del conocimiento* como un sistema, y específicamente como un sistema de conceptos. Los componentes del sistema son, por lo tanto, los conceptos de un dominio, expresados por designaciones unívocas (notaciones, epígrafes, términos preferentes, etc.), interconectados mediante distintos tipos de relaciones y estructurados en un orden determinado. En el área de la BDCI, la función de un SOC es facilitar la organización y recuperación del conocimiento registrado. Son ejemplos de SOC para la BDCI los sistemas de clasificación, las listas de epígrafes, las listas de autoridades, los tesauros, las taxonomías y las ontologías.

Lengua natural y organización del conocimiento

La lengua natural también se utiliza en el proceso de búsqueda, por ejemplo en aquellos sistemas de recuperación que permiten buscar por las palabras claves de título, de resumen o del texto completo del documento. No obstante, presenta algunas dificultades derivadas principalmente de la sinoni-

mia y la homonimia. Por ejemplo, si el autor de un documento utiliza el término ‘canes’, el usuario no podrá recuperarlo usando el término de búsqueda ‘perros’. Si el usuario busca el término ‘violeta’ por interés en la flor, también recuperará documentos sobre el color violeta o las personas de nombre Violeta. En otras palabras, el usuario final de un sistema de recuperación puede perder información debido a la sinonimia y obtener recuperaciones falsas a causa de la homonimia. Para solucionar estos problemas es necesario recurrir a un SOC porque su vocabulario es unívoco.

Durante largo tiempo se ha planteado la controversia respecto a la recuperación por lengua natural o por medio de un SOC como si fueran antagónicos, cuando en realidad son complementarios. Elaine Svenonius (1986) y Jennifer Rowley (1994) han distinguido en esto varias eras, a partir del siglo diecinueve.

La primera era comenzó con el bibliotecario británico Samson Low, quien en 1854 indizó el *British catalogue of books* utilizando palabras claves del título de los libros, es decir, la lengua natural usada por los autores. Su discípulo Andrea Crestadoro llevó a cabo una experiencia similar con el catálogo de la Manchester Free Library.

Por el contrario, cuando el estadounidense Charles A. Cutter publicó en 1876 las primeras reglas para redactar los epígrafes y Melvil Dewey hizo lo propio con su sistema de clasificación, dieron lugar a los primeros SOC. Hacia la primera mitad del siglo veinte la práctica habitual en las bibliotecas era indizar con una lista de epígrafes y clasificar con un sistema de clasificación. En esta era surgieron los SOC universales y precoordinados más conocidos: LCSH, Sears, SCDD, CDU, CC, BC.

La segunda era se inicia con la aparición, a partir de 1958, de los primeros índices realizados con una computadora: los famosos KWIC, KWOC y KWAC, elaborados mediante la extracción automática de las palabras claves del título de los documentos. Esta técnica abrió de nuevo el debate entre la lengua natural y los SOC. La controversia se vio reforzada aún más por los llamados estudios de Cranfield, llevados a cabo en Gran Bretaña entre 1957 y 1962. En particular los resultados de Cranfield 2 señalaban que la lengua natural, sin control o con un mínimo de control de vocabulario, lograba una recuperación tan eficiente como los SOC (Dubois, 1987; Lancaster, 2002; Rowley, 1994; Svenonius, 1986). Entonces, ¿para qué gastar dinero y es-

fuerzo en el diseño de un SOC y en la indización realizada por humanos, si basta con la lengua natural y una computadora programada para extraer las palabras claves?

La tercera era se inicia justamente con la proliferación de las bases de datos bibliográficas y los catálogos en línea, que permitieron llevar a cabo nuevas investigaciones sobre el desempeño de los usuarios en la búsqueda por materia, entre ellos el estudio realizado por Carol Tenopir (1985) sobre la base de datos *Harvard Business Review* de texto completo. Estas investigaciones demostraron que tanto la búsqueda por lengua natural como aquella por medio de un SOC ofrecen resultados efectivos, pero distintos. Para aclarar mejor esta idea, es necesario explicar primero las ventajas y desventajas que ofrecen tanto la lengua natural como los SOC para la recuperación, así como dos variables de medición que son las tasas de precisión y acierto.

En la Tabla 2 se comparan las ventajas y desventajas de la lengua natural y los SOC para la recuperación de información (Aitchison *et al.*, 2000; Dubois, 1987).

Tabla 2. Lengua natural vs. SOC

Lengua natural	Sistemas de organización del conocimiento		
Ventajas	<p>Mayor especificidad Mayor exhaustividad Actualización inmediata Evita errores del indizador</p> <p>El usuario utiliza sus propios términos Bajos costos en el almacenamiento Compatibilidad con otras bases de datos</p>	Desventajas	<p>Menor especificidad Menor exhaustividad Actualización mediata Permite errores del indizador</p> <p>El usuario tiene que aprender los términos Altos costos en el almacenamiento Incompatibilidad con otras bases de datos</p>
Desventajas	<p>Problemas de sinonimia y homonimia, singular y plural</p> <p>Se pierden conceptos implícitos Puede haber asociación incorrecta de términos</p>	Ventajas	<p>Control de sinónimos, homónimos y formas gramáticas</p> <p>Expresa conceptos implícitos Términos multipalabra normalizados</p>

Por otro lado, los experimentos de Cranfield introdujeron dos variables operativas para medir el desempeño de los sistemas de recuperación (Claverdon & Mills, 1985; Foskett, 1996; Lancaster, 2002). Las variables son:

- Tasa de precisión: es la proporción de documentos recuperados que son relevantes. Dicho de otra manera, es la capacidad de reducir o limitar una búsqueda, hasta recuperar si se necesita un solo recurso.
- Tasa de acierto: es la proporción de documentos relevantes que son recuperados, es decir, la capacidad de ampliar una búsqueda o de encontrar todos los recursos que tratan sobre un mismo tema.

Entonces, los estudios a los que hacíamos referencia, realizados sobre bases de datos de texto completo, demostraron que la lengua natural y los SOC pueden mejorar tanto la precisión como el acierto, pero de diferente manera (Dubois, 1987; Rowley, 1994; Svenonius, 1986; Tenopir, 1985):

- La lengua natural mejora la precisión por el mayor nivel de especificidad y la actualización de sus términos, mientras que los SOC lo hacen mediante el uso de frases o términos multipalabra y el control de homónimos.
- La lengua natural mejora el acierto cuando se realiza sobre un resumen o el texto completo por la exhaustividad de su cobertura. En cambio, los SOC mejoran el acierto por el control de sinónimos y las relaciones jerárquicas y asociativas.

La conclusión a la que se llegó es que la combinación de ambas modalidades en el mismo sistema de recuperación mejora la eficacia, porque las ventajas de una compensan las desventajas de la otra, es decir, la lengua natural y los SOC son complementarios y combinados ofrecen al usuario mayores oportunidades de lograr una búsqueda exitosa.

Además, otros investigadores demostraron que los usuarios finales tienen también sus preferencias, ya sea por la lengua natural o por los SOC (Ensor, 1992; Fidel, 1991, 1992).

Una nueva era se abre en el presente y hacia el futuro. Por un lado Google, con su algoritmo de búsqueda, ha revalorado la recuperación por lengua natural, mientras que la Web Semántica propuesta por el World Wide Web Consortium (W3C) promueve ciertos formatos comunes para los datos y la inclusión de con-

tenido semántico en las páginas Web, con la idea de permitir no solo la conexión a través de la red, sino también la reunión de conceptos (Dextre Clarke, 2008; W3C, 2011). Esta propuesta de la Web Semántica abre nuevos caminos para los SOC y ofrece varias herramientas específicas para su gestión, como los formatos SKOS para desarrollar SOC en línea y OWL para las ontologías.

Procesos de organización del conocimiento

El *Diccionario de la lengua española* define *proceso* como “conjunto de fases sucesivas de un fenómeno natural o de una operación artificial” (Real Academia Española, 2013). Otra fuente más específica propone la siguiente definición:

Secuencia de procedimientos interdependientes y vinculados que, en toda etapa, consume uno o más recursos (empleado, tiempo, energía, máquinas, dinero) para convertir entradas (datos, materiales, partes, etc.) en salidas. Estas salidas luego sirven de entrada para la siguiente etapa hasta lograr una meta conocida o un resultado final (Business Dictionary, 2013).

Así, por ejemplo, el proceso de la condensación o redacción de resúmenes consume recursos (condensador, computadora, electricidad, dinero) para transformar elementos de entrada (documento, normas, instructivos, etc.) en una salida o resultado —el resumen—, que a su vez puede ser la entrada de otro proceso como la indización y contribuir a un producto final como el registro bibliográfico de un catálogo o de una bibliografía.

Podemos afirmar, entonces, que los POC son aquellos procesos que posibilitan organizar el conocimiento con el fin de facilitar su recuperación. También nos permitimos considerar dos tipos de POC: a) los POC de la descripción de contenido y b) los POC del diseño y desarrollo de SOC.

Procesos de organización del conocimiento de la descripción de contenido

Nos referimos aquí a los procesos involucrados en la correlación concepto-objeto, descrita por Dahlberg (2006a) como una de las aplicaciones de la organización del conocimiento.

En 1949, la documentalista británica Edith Ditmas describió las tareas que se realizan en las bibliotecas y centros de documentación como un proceso exacto y continuo, que denominó *proceso documental*. Casi tres décadas después, el documentalista francés Jacques Chaumier renombró a este proceso *cadena documental*, entendiendo que en realidad se trata de numerosos procesos, cada uno de los cuales constituye un eslabón concatenado con los demás para culminar en uno o varios productos, de modo similar a lo que en el ámbito industrial se conoce como una cadena productiva (Chaumier, 2004).

Los procesos de la cadena documental reciben distintas designaciones según se trate de la bibliotecología o de la documentación y la ciencia de la información.

La norma ISO 5127, referida al vocabulario de la información y la documentación, ofrece dos definiciones de *biblioteca*:

“Biblioteca 1: colección organizada de documentos seleccionados y procesados para uso por una población destinataria predefinida. Biblioteca 2: es un organismo o parte de un organismo en el cual se construye la Biblioteca 1 y se hace disponible por los servicios de su personal” (ISO, 2001: 35).

De modo que una biblioteca requiere, obligatoriamente, la creación de una colección de documentos, real o virtual, que debe ser organizada. Los procesos que permiten organizar la colección se denominan en nuestro medio *procesos técnicos* y su producto final es el catálogo de la biblioteca, que consiste en una “lista ordenada de un sistema de recuperación bibliográfica, que permite el acceso a los datos y ubicaciones de los documentos en una o varias colecciones” (ISO, 2001: 72). Los procesos técnicos incluyen: a) el desarrollo de la colección, b) la catalogación descriptiva, c) la catalogación de materias y d) la condensación. Los POC son los dos últimos.

El proceso de la catalogación de materias abarca procedimientos tales como el análisis conceptual, la asignación de puntos de acceso verbales y controlados que representan el contenido del documento en un registro bibliográfico y la asignación de una signatura topográfica u otro elemento que remita a la ubicación física del documento.

El proceso de la condensación permite incluir en el registro bibliográfico una forma condensada como la nota de contenido o el resumen. Aunque no todas las bibliotecas incluyen estas formas condensadas en sus catálogos, su utilidad ha sido señalada por otros autores y recomendada por asociaciones internacionales (Dwyer, 1991; IFLA, 2011).

Por otro lado, un *centro de documentación* es un “organismo que lleva a cabo funciones de documentación” (ISO 2001, p. 36), es decir, las funciones de “compilación y procesamiento continuos y sistemáticos de la información registrada, con el propósito de almacenamiento, clasificación, recuperación, utilización, o transmisión (ISO 2001: 10). También se conocen como servicios de indización y resumen.

Un centro de documentación no necesita desarrollar una colección propia, como es el caso de una biblioteca. Si bien se realiza una selección de los documentos que han de procesarse, estos no tienen que pertenecer necesariamente al centro de documentación, ni tampoco formar una colección.

Según varios documentalistas españoles, los procesos de la cadena documental que se llevan a cabo en un centro de documentación se dividen en dos grandes etapas: a) la colecta o selección de documentos y b) el tratamiento de la información. A su vez, el tratamiento de la información se divide en otras dos etapas: el análisis documental y la recuperación (López Yepes, 2004; Pinto Molina, 2001; Ruiz Pérez, 1992). El análisis documental es entendido de la siguiente manera:

Consiste en extraer de un documento sus elementos más significativos al objeto de expresar su contenido de una forma distinta a la que presenta el original y por supuesto abreviada (Pinto Molina, 2001: 4).

Aquella operación del proceso documental que tras un reconocimiento o estudio intelectual y objetivo del documento, transforma la información en éste contenida, ofreciéndola en productos que hacen posible su adecuada identificación, selección, recuperación y utilización por parte del usuario interesado, dentro de un conjunto documental más amplio (Ruiz Pérez, 1992: 55).

Se reconocen dos tipos de análisis documental: el análisis documental

formal y el análisis documental de contenido, o simplemente análisis formal y análisis de contenido. El análisis de contenido incluye dos POC: la indización y la condensación (López Yepes, 2004; Pinto Molina, 2001; Ruiz Pérez, 1992).

Otros documentalistas vinculados a la Unesco usan una denominación distinta (Guinchat & Menou, 1990; Wersig & Neveling, 1976). Así, se prefieren los términos *descripción documental* y *descripción de contenido*, que se definen como:

Descripción de un documento que contiene todos los datos acerca de un documento que pueden ser de valor en un sistema de documentación específico; incluye por lo general la descripción bibliográfica y la descripción de contenido que surge del análisis documental, la condensación y la indización (Wersig & Neveling, 1976, p. 98).

Se llama descripción de contenido (DC) por analogía con la expresión “descripción bibliográfica” (DB), al conjunto de operaciones mediante las cuales se describe aquello de que trata un documento o un asunto (hecho, conceptos, números, imágenes, etc.) y los productos resultantes. Estas operaciones y productos se denominan comúnmente clasificación, indización, condensación y análisis (Guinchat & Menou, 1990: 121).

De modo que la descripción documental abarca la descripción bibliográfica y la descripción de contenido, y esta última incluye varios POC: la clasificación, la indización y la condensación.

El producto final por excelencia de un centro de documentación es una bibliografía especializada, que se define como un “sistema de recuperación bibliográfica que permite el acceso a los datos que describen e identifican documentos” (ISO, 2001: 72).

Ahora bien, ciertas bibliotecas —como las nacionales y algunas especializadas— también producen bibliografías y por lo tanto es correcto referirse al proceso de indización en bibliotecas.

En este sentido, algunos autores consideran que la catalogación y la indización son lo mismo y han establecido un paralelismo entre ambas. De este modo, se refieren a una indización descriptiva y una indización de materias

(Hjørland, 2007b; IFLA, 2011; Lancaster, 1995). También hay discrepancias en cuanto a considerar el proceso de clasificación como una forma de indización (IFLA, 2011; ISO, 2001; Lancaster, 2002), la indización como una forma de clasificación (Cleveland & Cleveland, 2013; Naumis Peña, 2003) o la indización y la clasificación como procesos distintos (Dahlberg, 2006a; Tudhope *et al.*, 2006). Aquí no profundizaremos en esta polémica.

Por otro lado, se reconocen dos tipos de indización (Foskett, 1996):

- La indización derivada que surge de la extracción automática de palabras claves del título, el resumen o el texto completo del documento. También se conoce como indización automática.
- La indización asignada que consiste justamente en la asignación, por un indizador, de un epígrafe, un término preferente o una notación.

A su vez, la indización puede ser precoordinada o poscoordinada. Es precoordinada cuando los conceptos son coordinados por el indizador en la etapa de almacenamiento. Por el contrario, la indización es poscoordinada cuando los conceptos son coordinados por el usuario final en la etapa de recuperación. El tesauro es un SOC que se aplica al proceso de la indización asignada y poscoordinada.

Los procesos anteriores, de una u otra manera, también se cumplen en otros sistemas de recuperación, por ejemplo en los sitios Web, donde los contenidos pueden indizarse y condensarse mediante registros metadatos que facilitan la recuperación.

Procesos de organización del conocimiento para el diseño y desarrollo de SOC

Son aquellos procesos que tienen que ver con lo que Dahlberg (2006a) denomina *construcción de sistemas de conceptos*. Incluyen, entre otros, la definición y análisis del dominio, el control del vocabulario, la determinación de las relaciones semánticas o de otra índole, la construcción del orden sistemático, la publicación, la evaluación y actualización continua del SOC. Se explicarán con mayor detalle en capítulos posteriores, aplicados al diseño y desarrollo del tesauro.

Autoevaluación

¿En qué categorías incluyen al tesauro los autores Hodge, Tudhope *et al.*, Zeng y Souza *et al.*?

¿Cuáles son las diferencias y semejanzas que muestran las herramientas semánticas mencionadas por Hjørland?

¿Cómo define sistema de organización del conocimiento?

¿Qué ventajas y desventajas tienen para la recuperación, respectivamente, el uso de la lengua natural y el uso de un sistema de organización del conocimiento como el tesauro?

¿Por qué la lengua natural y los sistemas de organización del conocimiento se complementan en la etapa de recuperación?

¿Qué procesos de organización del conocimiento de la cadena documental se llevan a cabo, generalmente, en una biblioteca?

¿Qué procesos de organización del conocimiento de la cadena documental se llevan a cabo, generalmente, en un centro de documentación?

Distintos enfoques para estudiar los SOC

En el capítulo 2 hicimos referencia a la diversidad terminológica empleada para designar a los SOC. En parte, tal situación se debe a que estos han sido estudiados desde distintos puntos de vista. A continuación analizaremos los SOC de acuerdo con tres enfoques: lingüístico, terminológico y bibliotecológico.

Enfoque lingüístico

La lingüística documental, muy difundida en las últimas décadas del siglo veinte, propone apropiarse de la teoría lingüística particularmente de la corriente estructuralista para fundamentar el diseño y aplicación de los SOC, que bajo este enfoque han sido designados *lenguajes documentales*. García Gutiérrez lo explica de la siguiente manera:

La denominación “lingüística documental” surge en el momento en que los elementos que intervienen en el proceso comunicativo-documental exceden su tradicional naturaleza lógica y mecanicista y son descritos, se transforman y organizan, al menos esa es nuestra propuesta, como elementos de comportamiento fundamentalmente lingüístico manteniendo y potenciando sus objetivos habituales [...] Se pretende demostrar, en realidad, que la exaltación del mensaje, de la comunicación, de los signos, de la estructura y de los enunciados en Documentación sugiere un mayor aprovechamiento de la Teoría Lingüística como fundamento de los métodos y procesos documentales (García Gutiérrez, 1990: 11).

La profesora cubana Dolores Vizcaya Alonso adhiere a esta posición y de cierta manera enfatiza el vínculo con la corriente estructuralista iniciada por Ferdinand de Saussure:

Todo lo que sobre lenguajes documentarios se estudie, se tendrá que hacer, de una forma u otra, dentro del campo de la lingüística; y cuando de esta disciplina se trata, no se puede comenzar a escribir nada sin dedicarle las primeras líneas a quien tan merecidamente fuera llamado el padre de la lingüística moderna: Ferdinand de Saussure (Vizcaya Alonso, 1997: 15).

Así como el objeto de estudio de la lingüística es la lengua natural, el objeto de estudio de la lingüística documental es el lenguaje documental, es decir los SOC.

Dentro de este enfoque haremos algunas precisiones. En primer lugar, podemos distinguir entre lenguaje y lengua. El *lenguaje* es una capacidad abstracta y general de comunicación. Existen diversas formas de comunicación humanas y no humanas que constituyen un lenguaje, pero no una lengua en sentido estricto; por ejemplo, el lenguaje corporal humano o el lenguaje de las abejas. En cambio la lengua, si bien es un tipo de lenguaje, es un sistema concreto y particular de signos que forman un léxico o vocabulario, una semántica y una sintaxis (ISO, 2001; Lyons, 1997; Naumis Peña, 2008).

Es importante aclarar que el inglés, un idioma que ejerce fuerte influencia sobre la terminología tanto de la lingüística como de la BDCI, posee un solo vocablo *language* para expresar tanto el concepto de lenguaje como el de lengua. Por el contrario, otras lenguas tienen un vocablo diferente para cada concepto: en francés *langage* y *langue*, en italiano *linguaggio* y *lingua*, en portugués *linguagem* y *língua*, en español *lenguaje* y *lengua* (ISO, 2001; Lyons, 1984; Naumis Peña, 2008). La traducción del inglés a estos últimos idiomas hizo que se acuñaran en la BDCI los términos *lenguaje natural*, *lenguaje artificial*, *lenguaje de indización*, *lenguaje de descriptores*, *lenguaje documental* o *lenguaje documentario*, es decir que se usara *lenguaje* en lugar de *lengua*. Aquí utilizaremos *lengua natural* para referirnos al objeto de estudio de la lingüística, pero debido a su uso generalizado aceptaremos *lenguaje documental*, *lenguaje documentario*, *lenguaje de indización* y otros términos similares usados en la BDCI, entendiendo que se trata de lenguas, en este caso artificiales.

Ahora es preciso distinguir entre lengua natural y lengua artificial. La lengua natural es aquella hablada por una comunidad lingüística particular. El

francés, el ruso, el mandarín, el swahili, el guaraní o el quekchí son ejemplos de lenguas naturales. En la actualidad, se hablan miles de lenguas naturales en todo el mundo y se conocen también numerosas lenguas muertas o extintas.

Una lengua natural tiene variadas funciones, ya que con ella se puede redactar un informe técnico, narrar una historia cierta o ficticia, dar una orden, discutir un asunto, influir sobre otras personas, componer un poema o establecer contacto con otros mediante un simple *hola*. Además, la lengua natural posee un léxico plurisemántico y su sintaxis se establece con el uso.

Por el contrario, las llamadas lenguas artificiales, entre las que se encuentran los lenguajes documentales, tienen una única función, su vocabulario es unívoco y su sintaxis se establece con anterioridad al uso. Entonces, los lenguajes documentales o SOC son lenguas artificiales, cuya principal función es organizar el conocimiento socializado y registrado para que pueda ser recuperado. Además, acorde con lo que señala Svenonius (2003), los SOC como lenguas artificiales constan también de un léxico o vocabulario, una semántica y una sintaxis.

El léxico o vocabulario es el conjunto finito de los signos de una lengua, ya sea en general o pertenecientes al uso de una región, una actividad determinada, un campo semántico dado, una persona o un grupo de personas (Barité *et al.*, 2013; ISO, 2001; López Yepes, 2004; Real Academia Española, 2011; Wersig & Neveling, 1976).

El vocabulario de un SOC es por lo tanto el conjunto de signos que han sido aceptados para su aplicación en los procesos de organización del conocimiento. Al igual que en la lengua natural, los signos de un SOC constan de significado (imagen mental) y de significante (imagen acústica o, en el caso de los SOC, imagen gráfica) y la relación entre significado y significante es arbitraria.

Los signos de un SOC pueden tomarse de la lengua natural, como sucede en las listas de epígrafes y en los tesauros, o de otro tipo de sistema, como es el caso de las notaciones de un sistema de clasificación, para las cuales se utilizan números, letras o una combinación de ambos. En todos los casos y en virtud de un artificio que se conoce como *control del vocabulario*, cada signo del SOC es despojado de todos sus significados menos uno y de todos sus significantes menos uno, de modo que la totalidad de los signos resultan unívocos, eliminándose así la sinonimia y la homonimia que caracterizan la

lengua natural, pero que presentan tantos problemas para la recuperación.

La profesora estadounidense Marcia Bates (1988) señala que el control del vocabulario también incluye las reglas para formalizar el significante (forma gramatical, género y número, estructura del sintagma) y los británicos Aitchison *et al.* (2000) afirman que el control del vocabulario consta de cuatro fases:

- Control del significante: elección de un solo significante, su forma gramatical, ortografía, género y número, etc.
- Control de sinónimos: selección de solo uno de los sinónimos disponibles para expresar un concepto.
- Control de signos especiales: selección de términos prestados, jergas, nombres comerciales y nombres propios.
- Control del significado o desambiguación: elección de un solo significado, que debe ser el más efectivo para los propósitos del SOC. Esta restricción se logra mediante calificadores, notas de definición y notas de alcance, como se verá más adelante en el capítulo 7.

El producto del control del vocabulario es un *vocabulario controlado*, que definimos como el conjunto restringido de signos de un SOC autorizados para uso en los POC de descripción de contenido (ISO, 2001; López Yepes, 2004).

Algunos autores consideran que *vocabulario controlado* es un sinónimo de lenguaje documental o de SOC y, por lo tanto, se afirma que los tesauros, las listas de epígrafes y los sistemas de clasificación son vocabularios controlados (ISO, 2011; Lancaster, 2002; NISO, 2010). Nosotros no estamos de acuerdo con esto porque consideramos que el vocabulario controlado es un componente del SOC, junto con la semántica y la sintaxis.

La semántica es “la rama de la Lingüística que se ocupa del sentido o el significado de los signos, así como de la relación entre los mismos, tanto desde un punto de vista sincrónico como diacrónico” (Barité *et al.*, 2013). En un SOC, la semántica se evidencia en la “red de significados que se establece en un sistema de organización del conocimiento, entre los distintos términos que forman su estructura” (Barité *et al.*, 2013). En un tesauro, esta red de significados se explicita en las relaciones de equivalencia, jerarquía y asociación (Hjørland, 2007a,b; ISO, 2001; Svenonius, 2003; Wersig & Neveling, 1976).

Finalmente, la sintaxis es la “disciplina que se ocupa del estudio de la estructura gramatical de la lengua, y de las reglas que gobiernan la construcción de frases o sentencias que forman el discurso, así como de la relación entre las palabras que las forman” (Barité *et al.*, 2013). En un SOC, la sintaxis es el “conjunto de reglas utilizadas para combinar descriptores o símbolos de clasificación, de modo tal que expresen una secuencia lógica” (Barité *et al.*, 2013). En este último sentido, la sintaxis de los SOC puede ser pre-coordinada o poscoordinada (Foskett, 1996; Svenonious, 2003; Wersig & Neveling, 1976).

Barité *et al.* ofrecen dos acepciones para el término *precoordinación*:

a. “Procedimiento de combinación de dos o más términos o símbolos notacionales incorporado originalmente en un lenguaje documental” (Barité *et al.*, 2013). En este caso, el signo del SOC puede ser una frase que combine los conceptos que se desea representar (Miller & Teitelbaum, 2002). Ejemplos de este tipo de precoordinación son los siguientes epígrafes de LEMB:

comunicación en administración de bibliotecas
conservación y restauración de libros
programas nacionales para adquisición y catalogación

b. “Combinación de términos o símbolos que realiza el indizador o el clasificador considerando las temáticas involucradas en el documento que está procesando” (Barité *et al.*, 2013). En este caso se debe construir una cadena precoordinada cuyos componentes deben seguir un orden de mención preestablecido (Miller & Teitelbaum, 2002). Los siguientes epígrafes de LEMB y notaciones de CDU son ejemplo de cadenas precoordinadas:

bibliotecarios-ética profesional
bibliotecas-historia-hasta 400 d.C.
bibliotecología-historia-Argentina-1810-1902
004:023
026(82)”20”

Las listas de epígrafes como LCSH, LEMB, Sears o ARMARC y los sistemas de clasificación como SCDD y CDU son precoordinados.

Por otro lado, también Barité *et al.* (2013) definen *poscoordinación* como “procedimiento de libre combinación de términos o descriptores al momento en que se define la ecuación de búsqueda”. En este caso, los términos se almacenan en cualquier orden y es el usuario final quien los coordina en su estrategia de búsqueda (Miller & Teitelbaum, 2002). Los siguientes son ejemplos tomados del Tesouro Spines:

economía
economía humana
economía de mercado

Así, los tesauros son los SOC apropiados para la indización poscoordinada.

Enfoque terminológico

Como se verá en capítulos posteriores, este enfoque terminológico es de primordial relevancia para el diseño y desarrollo de un tesouro.

Comenzaremos por aclarar que la palabra *terminología* puede entenderse en dos sentidos. Por un lado, la terminología (con minúscula) es el conjunto de términos especializados propios de una disciplina o un campo de conocimiento. De este modo, es común referirse a la terminología médica, la terminología informática, la terminología jurídica, etc. También se conoce como *lengua de especialidad*.

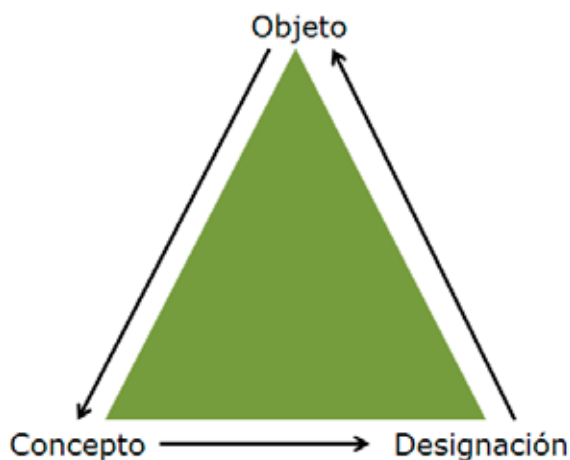
Por otro lado, la Terminología (con mayúscula) es la disciplina o el campo multidisciplinario que estudia las unidades terminológicas y los productos terminológicos (glosarios, diccionarios, etc.) y propone una teoría sobre ellos.

Quienes estudian los SOC desde el punto de vista de la Terminología, lo hacen generalmente sobre la base de la tríada objeto-concepto-designación, que resulta de especial utilidad cuando se aplica al diseño de un tesouro (Figura 1).

La norma ISO 25964 para el diseño y desarrollo de tesauros afirma que un *término* es “una palabra o frase usada para etiquetar un concepto”; a su vez, define *concepto* como “unidad de pensamiento” (ISO, 2011: 3, 12). Si bien estas afirmaciones son correctas, llama la atención la escasa y superficial

información que ofrece la norma tanto de *concepto* como de *término*, sobre todo porque esta distinción es indispensable para comprender el modelo de datos recomendado por la norma, que se describe en el capítulo 12 de la presente obra.

Figura 1. Triángulo del concepto. El concepto es la representación mental de un objeto y se expresa mediante una designación que denota al objeto



Barité afirma que los profesionales de la BDCI hemos sido extremadamente descuidados en el estudio de los conceptos, tendiendo a “trabajar apenas con las etiquetas o 'fachadas' de los conceptos, sin considerar los aspectos relativos a la construcción y definición de los mismos” (Barité, 2000a: 31).

Debido a la importancia que esto tiene para el diseño de un tesoro, recurriremos a otra fuente para explicarlo con mayor detalle: nos referimos a la norma ISO 704:2009 sobre principios y métodos del trabajo terminológico (ISO, 2009). Esta norma terminológica tiene numerosos puntos en común con las normas para el diseño de tesauros, pero ofrece precisiones que no se encuentran en estas últimas.

Objeto

Para la norma ISO 704:2009, un objeto es cualquier cosa que un ser hu-

mano pueda percibir o concebir. Dahlberg (1978) lo llama *referente*.

El objeto puede ser concreto o material, como un edificio, una silla, un libro, un niño, un bosque. También puede ser abstracto o inmaterial, como la felicidad, la tristeza, el pensamiento, las ideologías, las acciones. Por último, el objeto puede ser imaginario como una sirena, un centauro, Hamlet, la Hogwarts School o Macondo.

Todos los objetos tienen una o más propiedades que los caracterizan y permiten comprenderlos y aprehenderlos.

Concepto

Un concepto es la representación mental de un objeto. A través de la observación y de un proceso de abstracción llamado *conceptualización*, un objeto del mundo real o del mundo ideal es categorizado como una unidad de conocimiento que es precisamente el concepto (ISO, 2009). Otros autores lo llaman *noción* (Barité, 2000a), pero aquí usamos *concepto*, siguiendo la citada norma ISO 704:2009.

Barité (2000b: 40) identifica los siguientes rasgos definitorios de *concepto*:

- a. Todo concepto es una construcción abstracta, que utiliza el lenguaje como vía para representar (o sea, volver a presentar) de un modo comprensible o inteligible, un objeto o un fenómeno de la realidad (presente, pasada o incluso futura) o del reino de lo ideal.
- b. Todo concepto o noción corresponde a la expresión de una unidad aislable, autónoma de pensamiento. Es, para decirlo de otra manera, una parte significativa de la realidad material o ideal sobre la cual la reflexión humana ejerce una segmentación o delimitación convencional [...].
- c. Asimismo, y por consecuencia, cada concepto o noción se constituye en una unidad mínima y autosuficiente de conocimiento.
- d. La construcción de un concepto implica necesariamente su definición, usualmente por medio de claves de lenguaje [...].
- e. Desde una óptica terminológica, todavía es posible agregar un quinto rasgo o atributo de las nociones, y es que cada una de ellas, en cuanto unidad orgánica de conocimiento, se inserta necesariamente en una rama, una parte, una disciplina, en suma, del saber especializado. No cabe ha-

blar de conceptos en Terminología sin el marco referencial de una disciplina (científica, técnica) o de una especialidad, y del sistema de conceptos que la representa. Ausente ese marco, solo encontramos palabras de la lengua general con sus significados, pero no conceptos especializados, pues los conceptos son en última instancia, unidades cognitivas que adquieren su significación plena solo en el contexto de un sistema nocional disciplinario.

Un concepto puede ser general o individual (ISO, 2009). Es un concepto general cuando representa a un conjunto de objetos que comparten propiedades comunes, por ejemplo: ‘organismos internacionales’, ‘universidades’, ‘países americanos’, ‘naves espaciales’. En cambio, es un concepto individual cuando representa un objeto único, distinto de todos los demás y que constituye una unidad inconfundible, por ejemplo: ‘Organización de las Naciones Unidas’, ‘Universidad de Buenos Aires’, ‘Paraguay’, ‘Apolo 11’.

El concepto, ya sea general o individual, se construye a partir de las características que predicen o declaran las propiedades del objeto. Cada propiedad del objeto se abstrae como una característica del concepto. Entonces, un concepto es la abstracción de un objeto y una característica es la abstracción de una propiedad del objeto.

Las características pueden ser, básicamente, de dos tipos:

- Características esenciales: son aquellas características necesarias y suficientes para determinar un concepto y son imprescindibles para definir el objeto sin ambigüedad. Por ejemplo, el concepto ‘caballo’ tiene, entre otras, las siguientes características esenciales: es un mamífero perisodáctilo, pertenece al género *equus*, especie *ferus*, subespecie *caballus*, es herbívoro.
- Características accidentales: son aquellas características adicionales o posibles que solo se pueden generalizar respecto a una parte del objeto o a ciertos tipos de objetos. Siguiendo con el ejemplo del concepto ‘caballo’, las siguientes serían características accidentales: es un alazán, posee una marca blanca en la cara de tipo lucero, tiene calcetines blancos en tres patas, compete en deportes ecuestres.

El conocimiento de las características del concepto hace posible: a) cons-

truir el concepto, b) definir el concepto, c) establecer las relaciones entre los conceptos y d) modelar un sistema de conceptos (Dahlberg, 1978).

Dado que las características y los conceptos son representaciones mentales o abstracciones, para comunicarlos es necesario recurrir a la lengua. Así, una característica se expresa mediante un enunciado y un concepto se expresa por medio de una definición o de una designación.

Definición

De acuerdo con Barité *et al.*, una definición es:

Conjunto de enunciados que describen un objeto (cosa, ser, fenómeno, entidad), en un nivel de abstracción dado, con la finalidad de establecer su identidad, y diferenciarlo inequívocamente de otros objetos. La definición de un término es un presupuesto básico para la comunicación, el intercambio en uno o varios idiomas y la argumentación, y constituye un elemento insustituible en la construcción del pensamiento científico. Debe cumplir, entre otros, los siguientes requisitos: precisión, concisión, claridad, autosuficiencia y un nivel de exhaustividad adecuado a los destinatarios (Barité *et al.*, 2013).

La definición establece los límites del concepto, hace posible su comprensión y lo distingue de otros conceptos. El conjunto de características que forma el concepto se denomina la *intensión* del concepto. El conjunto de objetos conceptualizados como un concepto se conoce como la *extensión* del concepto (ISO, 2009). Entonces, se puede establecer una definición por intención o una definición por extensión. Por ejemplo:

- Definición por intención de *caballo*: mamífero, herbívoro, perisodáctilo del género *Equus*, especie *ferus*, subespecie *caballus*. Posee cuello y cola poblados de cerdas largas y abundantes; domesticado desde hace unos 5500 años. En general es de gran porte, aunque varía en tamaño, pelaje, marcas y actividades a las que se destina.
- Definición por extensión de *caballo*: el caballo puede ser árabe, anglo árabe, criollo argentino, connemara, hannoveriano, poni australiano, holstein, poni camargue, appaloosa, haflinger, poni dartmoor, jaca galeasa, andaluz, de silla, poni new forst, lipizzano, de tiro holandés [...].

La norma ISO 704:2009 considera que el tipo de definición más conveniente en el área de la Terminología es la llamada *definición por intensión*. Este tipo de definición se compone de un concepto superordenado, que ubica el concepto en su contexto, seguido por las características que lo delimitan. Como señala Barité (2000a), el principio involucrado en la definición por intensión es el de género próximo y diferencia específica, que ha sido también señalado por otros autores para las definiciones en los tesauros y otros SOC (Aguado *et al.*, 1999; Hudon, 1996; Vickery, 1960a).

A continuación se presentan tres ejemplos de definición por intensión:

caballo

DEF mamífero del orden de los Perisodáctilos, solípedo, del género *Equus*, especie *ferus*, subespecie *caballus*. Es herbívoro, de cuello y cola poblados de cerdas largas y abundantes; domesticado desde hace unos 5500 años. Suele ser de gran porte, aunque varía en tamaño, pelaje, marcas y actividades a las que se destina.

caballo deportivo

DEF caballo que se destina a la práctica de los deportes ecuestres como el salto de obstáculos, la hípica, el polo, la doma, etc.

Secretariat

DEF caballo deportivo que se destacó en la hípica. Nació en Estados Unidos el 30 de marzo de 1970 y murió el 4 de octubre de 1989. Hijo de Bold Ruller en la yegua Somethingroval. Propiedad del stud Meadow Stable de Penny (Chenery) Tweedy. Participó en 21 carreras con una marca de 21-16-3-1. En 1973 ganó la Triple Corona y los tres años que corrió, 1972, 1973 y 1974 fue designado Caballo del Año.

El primer concepto (caballo) es el más general, mientras que el último (Secretariat) es el más específico. El género próximo se hereda de una definición a otra: caballo deportivo hereda el género caballo, mientras que Secretariat hereda el género caballo deportivo. Esta cualidad de heredar el género próximo resulta muy útil en un tesoro para establecer las relaciones jerárquicas, que a veces se designan como relaciones padre-hijo. Por ejemplo, de las tres defini-

ciones anteriores se puede determinar la relación jerárquica entre los conceptos:

caballo

caballo deportivo

Secretariat

Designación

La designación es la expresión de un concepto por medios verbales o simbólicos. Así, la moneda argentina puede designarse por medio del término *pesos* o del símbolo \$, mientras que el símbolo registrado legalmente para identificar de manera exclusiva uno o varios productos comerciales se puede designar con el término marca registrada o con el símbolo ®.

Entre las designaciones verbales se puede hacer una distinción entre término y nombre propio. Un concepto general se designa mediante un término, mientras que un concepto individual se designa mediante un nombre propio. Volviendo al ejemplo del apartado anterior, *caballo* y *caballo deportivo* son términos que designan conceptos generales. En cambio *Secretariat* es un nombre propio que designa un concepto individual.

En la lengua natural, un concepto puede ser expresado por una o varias designaciones (sinonimia). A su vez una designación puede expresar uno o varios conceptos (homonimia). Por el contrario, en un SOC cada concepto debe expresarse por una y solo una designación e inversamente cada designación debe expresar uno y solo un concepto. Según el tipo de SOC, la designación puede ser, por ejemplo, un epígrafe, una notación, un término preferente, o una forma autorizada del nombre.

Relaciones semánticas y sistemas de conceptos

Cada concepto es una unidad de conocimiento, pero no es una unidad aislada, sino que mantiene relaciones semánticas con otros conceptos. Estas relaciones permiten construir una estructura que constituye un sistema de conceptos, definido como “un conjunto de conceptos estructurado de acuerdo con las relaciones entre ellos” (ISO, 2009: 8).

Un SOC, en cuanto sistema de conceptos, debe definir los conceptos, designarlos en forma unívoca, relacionarlos por su semántica y organizarlos en un orden determinado. En el tesoro, las relaciones semánticas son de tres tipos:

- Relación de equivalencia: se establece entre un término o un nombre

propio elegidos para la indización y sus sinónimos o formas variantes.

- Relación jerárquica: se establece entre un concepto que representa un género o un todo y otro que representa una especie, una parte, o un ejemplo del anterior.
- Relación asociativa: se establece entre dos conceptos, que no pueden considerarse ni equivalentes, ni jerárquicos.

3.3 Enfoque bibliotecológico

Los *Functional requirements for subject authority data* (FRSAD) son parte del nuevo modelo conceptual que ha propuesto IFLA para las bases de datos bibliográficas y de autoridades, que se integra en la llamada familia FRBR con los requisitos funcionales para registros bibliográficos (FRBR) y los requisitos funcionales para datos de autoridad de nombres propios (FRAD) (IFLA, 2004, 2009b, 2010).

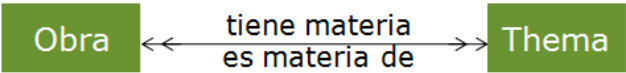
Originalmente, el modelo FRBR definió tres grupos de entidades según se muestra en la Tabla 1. Sin embargo, desde el primer momento se notó que las entidades del Grupo 3 no eran del todo correctas, ya que —como hemos visto— un concepto representa un objeto y este a su vez puede ser cualquiera de las entidades de los grupos 1, 2, 3 u otras más. De modo que para no alterar el modelo, los responsables de FRSAD se vieron obligados a postular dos superclases que responden a los términos en latín *thema* y *nomen*.

Tabla 1. Entidades del modelo FRBR-FRAD-FRSAD

Grupo de entidades	Entidad	Thema Nomen
1 FRBR	Obra Expresión Manifestación Ítem	
2 FRAD	Persona Familia Entidad corporativa	
3 FRSAD	Concepto Objeto Acontecimiento Lugar	

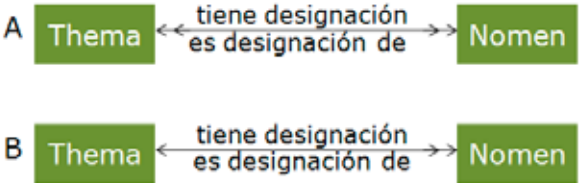
Thema (en plural *themata* o *themas*) es cualquier entidad usada como materia de una obra. Una obra puede tener como materia uno o varios *themas* y, a su vez, un *thema* puede ser materia de una o varias obras, según se muestra en la Figura 2.

Figura 2. Relación obra-*thema*. Nótese la doble flecha que relaciona estas entidades: una obra puede tener uno o más *themas* como materia y un *thema* puede ser materia de una o más obras.



Por otra parte, *nomen* (en plural *nomina* o *nomens*) es cualquier signo o secuencia de signos por los que se conoce, refiere o dirige un *thema*. Por ejemplo, el *thema* \$ puede tener como *nomens* anteojos, espejuelos, gafas o la notación de CDU 617.7-089.24.

Figura 3. Relación *thema*-*nomen*. A) En la lengua natural la flecha es doble en ambas direcciones, pues un *thema* puede tener uno o más *nomens* y a la inversa, un *nomen* puede designar uno o más *themas*. B) En un SOC la flecha es doble en un sentido y simple en el otro, porque un *thema* puede tener uno o más *nomens* (solo uno preferente), pero un *nomen* puede designar un solo *thema*.



La relación entre *thema* y *nomen* se esquematiza en la Figura 3. Como se puede observar, en la lengua natural, un *thema* puede ser designado por uno o

varios *nomens* (sinonimia) y a su vez un *nomen* puede designar uno o varios *themas* (homonimia). Al contrario, en un SOC un *thema* puede ser designado por uno o varios *nomens* (uno preferente y otros equivalentes o no preferentes), pero un *nomen* solo puede designar un *thema*.

Este enfoque bibliotecológico permite generalizar el asunto o materia más allá de las diferencias que existen entre los distintos SOC. Así, por ejemplo, el *thema* en un tesoro siempre es un concepto y, por lo tanto, el enfoque terminológico se ajusta perfectamente. En cambio en los SOC precoordinados, el *thema* puede ser un concepto o una combinación de conceptos.

El *nomen* en un tesoro puede ser un término preferente, un término no preferente, una forma autorizada del nombre o una forma variante del nombre, como se detalla en el capítulo 7.

En una lista de epígrafes el *nomen* puede ser un epígrafe, una forma autorizada del nombre, una cadena precoordinada o las formas variantes de todos estos. De forma similar, en un sistema de clasificación el *nomen* puede ser una notación o una cadena precoordinada de notaciones, así como los equivalentes verbales de estas. Por ejemplo, observemos los siguientes *nomens* de la lista de epígrafes ARMARC y del sistema de clasificación CDU:

iglesia y problemas sociales
enciclopedias y diccionarios para niños
oceanografía-aparatos e instrumentos
331.2 Salarios. Sueldos. Remuneraciones. Pagos
37:61(82) Enseñanza de la medicina en Argentina

Autoevaluación

¿Cuáles son los distintos enfoques con que se han estudiado los SOC?

Explique la diferencia entre lengua natural y lengua artificial.

¿Qué características posee el vocabulario de un SOC?

Indique cómo se evidencian la semántica y la sintaxis en un SOC.

Explique la tríada objeto-concepto-designación.

¿Qué tipo de definición se recomienda para un tesaurus?

¿Qué designación verbal le corresponde a un concepto general?

¿Qué designación verbal le corresponde a un concepto individual?

¿Qué significado tienen los términos *thema* y *nomen*?

Explique las relaciones entre obra y *thema* y entre *thema* y *nomen*.

Mencione las relaciones *thema-thema* y *nomen-nomen*.

El tesoro

En los capítulos anteriores hemos establecido un marco de referencia general. Ahora nos introduciremos específicamente en el tema del tesoro. Lo definiremos y caracterizaremos, analizando también su función y su tipología.

Definición y caracterización del tesoro

En el capítulo 2 se estableció que el tesoro es un SOC y que diversos especialistas lo han clasificado en distintas categorías: es una lista de relaciones (Hodge, 2000), un SOC basado en relaciones (Tudhope *et al.*, 2006), un modelo de relaciones (Zeng, 2008), una estructura de conceptos y relaciones (Souza *et al.*, 2010; 2012), o una herramienta semántica (Broughton *et al.*, 2005; Hjørland, 2007a). Estas categorías resaltan principalmente las relaciones entre los conceptos del tesoro.

También establecimos en el capítulo 3 que los SOC, en cuanto lenguas artificiales, constan de un vocabulario, una semántica y una sintaxis (Svenonius, 2003). Explicamos además la tríada objeto-concepto-designación, de importancia crítica para el tesoro (Barité, 2000a; ISO, 2009).

Sobre esta base y teniendo en cuenta las numerosas definiciones de *tesoro* que aparecen en la bibliografía, nos permitimos establecer nuestra propia definición como marco de referencia para esta obra:

- Un tesoro es un sistema de organización del conocimiento.
- Se compone de: a) un vocabulario controlado, b) una estructura de relaciones semánticas explícitas de equivalencia, jerarquía y asociación, c) un orden sistemático que evidencia el sistema de conceptos de un dominio, d) una sintaxis poscoordinada y e) una cobertura limitada a un dominio determinado.
- Se aplica a la organización y recuperación del conocimiento registrado en bases de datos bibliográficas, bibliotecas virtuales, portales Web, etc.

Como todos los SOC, un tesoro impone una cierta concepción del mundo, de acuerdo con el dominio que trata, los usuarios a los que está dirigido, la lengua en que se elabora, la cultura de una región, país, grupo étnico u organismo del que procede.

Componentes del tesoro

Hemos mencionado en nuestra definición los componentes del tesoro. A continuación explicaremos brevemente cada uno de ellos.

El vocabulario controlado

El tesoro, en cuanto SOC, consta de un vocabulario controlado formado por un conjunto de términos y nombres propios que designan, respectivamente, conceptos generales e individuales. En el vocabulario de un tesoro se elimina toda sinonimia y homonimia, con el fin de aportar consistencia a la recuperación.

Si bien las normas y los manuales para el diseño y desarrollo de tesoros incluyen recomendaciones respecto a la redacción de los términos y nombres propios según se detalla en el capítulo 7 de esta obra, es importante tener en cuenta que el elemento esencial del vocabulario de un tesoro es el concepto y no el término o el nombre que lo designa, ya que estos últimos pueden variar, pero no el concepto que debe definirse con precisión.

Las relaciones del tesoro

Los conceptos incluidos en un tesoro se vinculan semánticamente entre sí mediante relaciones de equivalencia, jerarquía y asociación.

La relación de equivalencia es la que se establece entre un término que ha sido seleccionado para el tesoro y los sinónimos que han sido descartados, o bien entre una forma aceptada del nombre propio y las variantes que este pudiera tener.

La relación jerárquica se establece entre un concepto que representa el género o el todo y otro que representa una especie, una parte o un ejemplo del anterior.

La relación asociativa es aquella que se establece entre dos conceptos, pero que no es ni equivalente ni jerárquica.

Un tesoro debe presentar las tres relaciones de modo explícito y recípro-

co; si carece de alguna de ellas no es un tesoro. Estas relaciones se describen con detalle en los capítulos 7 y 8.

Orden sistemático

Los conceptos del tesoro se ordenan formando un orden sistemático o clasificado. El orden sistemático es una característica del tesoro que hace evidente el sistema de conceptos del dominio y sitúa cada concepto en ese contexto, de modo similar al esquema de un sistema de clasificación.

La norma ISO 25964 recomienda tres posibles formas de construir el orden sistemático: a) por disciplinas o campos temáticos, b) por facetas y c) por facetas bajo disciplinas o campos temáticos. Se describen en el capítulo 9 de la presente obra.

Sintaxis poscoordinada

El tesoro es un SOC desarrollado especialmente para la indización poscoordinada, es decir, los términos y los nombres propios se almacenan en el registro bibliográfico o metadatos sin un orden de cita o de mención predefinido y son coordinados por el usuario en el momento de la recuperación. Esta característica también establece una diferencia entre el tesoro y las listas de epígrafes que son precoordinadas.

El dominio del tesoro

En general, se considera que el tesoro es un tipo de SOC limitado a una cierta especialidad, campo temático o dominio, distinto de otros SOC que procuran abarcar la universalidad del conocimiento. En efecto, el tesoro nació con el propósito de resolver problemas de recuperación en áreas específicas de la ciencia y la tecnología y, hasta ahora, los numerosos tesoros elaborados son especializados.

Sin embargo, no podemos evitar que nos surja una duda: ¿la especialización es realmente una característica del tesoro o no se ha presentado la oportunidad de desarrollar un tesoro universal?

Por un lado, desarrollar un tesoro es una iniciativa costosa; por otro lado, existen SOC universales que con periódicas actualizaciones han dado solución a las bibliotecas generales o similares. Entonces, es posible que no se haya planteado la necesidad de construir un tesoro general o enciclopédi-

co, lo cual no significa que no se pueda hacer. De hecho, las experiencias de interoperabilidad entre tesauros que se han llevado a cabo desde hace varias décadas, de alguna manera reflejan la necesidad de contar con tesauros más generales en cuanto a su alcance temático (Martínez Tamayo *et al.*, 2011b).

Más aún, la pretendida tesauroización de algunas listas de epígrafes universales como LCSH, Sears y, en América Latina, la versión digital de LEMB y la lista ARMARC, si bien insatisfactoria en sus resultados (Dykstra, 1988; Martínez Tamayo & Valdez, 2009; Rolland-Thomas, 1989), también sugiere la posibilidad de un tesoro universal.

No obstante y a pesar de nuestras dudas, aquí aceptamos el carácter especializado que suele asignarse al tesoro.

Las funciones del tesoro

El tesoro surgió con el fin de facilitar la indización y la búsqueda en sistemas de almacenamiento y recuperación (SAR) como los catálogos de biblioteca o las bibliografías, es decir, para “guiar al indizador y al usuario final en la elección del mismo término para el mismo concepto” (ISO, 2011: 15). Esta función por supuesto sigue vigente hoy en día. Sin embargo, a partir del surgimiento de la Web —y especialmente de la Web semántica— el tesoro también se ha destinado a otras funciones, como la navegación, la búsqueda expandida o la asistencia en la indización automática.

El tesoro en los registros bibliográficos y metadatos

En una base de datos bibliográfica como un catálogo de biblioteca o una bibliografía, el tesoro cumple su función tanto en el proceso de indización como en el de búsqueda, que están íntimamente vinculados, como las dos caras de una moneda. El profesor australiano Antony C. Fosskett (1996) explica la relación entre estos procesos basándose en el conocido modelo de la comunicación de Shannon-Weaver (Figura 4.1).

Brevemente, para Fosskett el autor del contenido de un documento emite un mensaje (la obra), en un determinado código. Luego, el indizador decodifica el mensaje del autor, lo interpreta y describe en un registro bibliográfico que es almacenado en la base de datos, emitiendo así un nuevo mensaje, en un nuevo código: los términos y nombres propios del tesoro (Figuras 1 y 2).

Figura 1. Modelo de los procesos de indización y recuperación. Muestra al autor como emisor, la obra como mensaje, el usuario final como receptor y al indizador como intérprete.

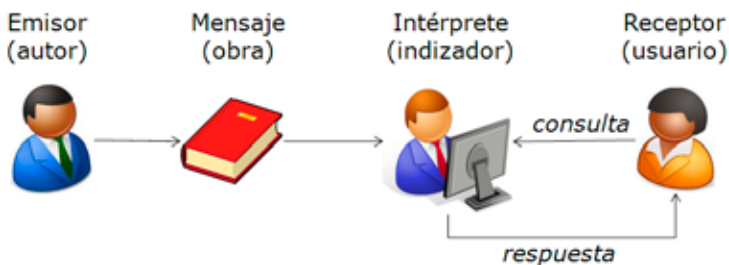
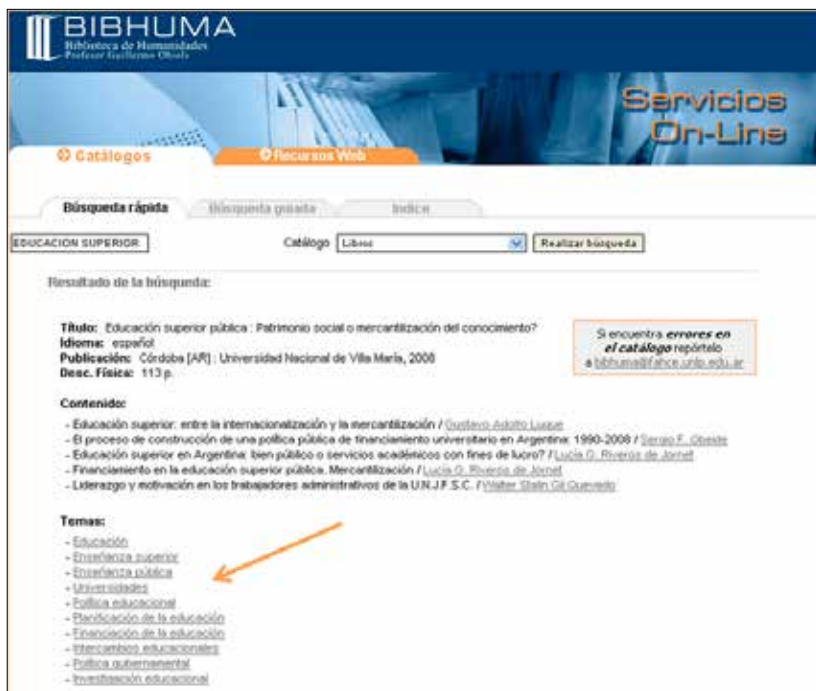


Figura 2. Registro bibliográfico de un catálogo en línea.¹ La flecha señala los términos preferentes, que han sido tomados del Tesauro de la Unesco.



¹ Registro bibliográfico del catálogo de la Biblioteca de la Facultad de Humanidades y Ciencias de la Educación de la Universidad Nacional de La Plata.

Por su parte, el usuario final presenta su consulta utilizando también los términos y nombres propios del tesoro y el sistema le devuelve una respuesta, que decodifica e interpreta; es, por tanto, el receptor final del mensaje.

El tesoro también puede usarse para la indización de un recurso Web mediante un registro metadatos. Este tipo de registros puede incrustarse en el mismo documento o puede generarse aparte y enlazarse por un hipervínculo (Angelozi & Martín, 2010). Su normalización se basa en un esquema metadatos como el de la Dublin Core Metadata Initiative (DCMI), dentro de cuyos elementos se encuentran las palabras claves, notaciones, epígrafes o términos preferentes de un tesoro, que es la opción recomendada (DCMI, 2012). En la Figura 3 se muestra un ejemplo de registro metadatos.

Figura 3. Registro de metadatos de un repositorio institucional.² Nótese las etiquetas DC que corresponden al esquema de la DCMI.

```
<META NAME="Author" CONTENT="Gamba, Viviana Lis">
<META NAME="Keywords" CONTENT="Bibliotecología y ciencia de la información">
<META NAME="citation_date" CONTENT="2006">
<META NAME="DC.date" CONTENT="2006">
<META NAME="citation_dissertation_institution" CONTENT="Universidad Nacional de La Plata.
Facultad de Humanidades y Ciencias de la Educación">
<META NAME="citation_title" CONTENT="Metodología para la descripción de contenidos de
imágenes fotográficas : ">
<META NAME="citation_keywords" CONTENT="Bibliotecología y ciencia de la información">
<META NAME="DC.subject" CONTENT="Bibliotecología y ciencia de la información">
<META NAME="citation_language" CONTENT="es">
<META NAME="citation_authors" CONTENT="Gamba, Viviana Lis">
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" >
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" >
<META NAME="DC.creator" CONTENT="Gamba, Viviana Lis">
<META NAME="DC.language" CONTENT="es">
<META NAME="DC.rights" CONTENT="This document is licensed under a Creative Commons License
Attribution-Noncommercial-No Derivative Works 2.5 Argentina. Authorization for web
publishing was granted to Humanities and Education Sciences Faculty, National University of
La Plata, by the author. This record was made at the Prof. Guillermo Obiols Library of the
Humanities and Education Sciences Faculty, National University of La Plata, Argentine."
xml:lang="en" >
<META NAME="DC.rights" CONTENT="Este documento se encuentra bajo una licencia Atribución-No
comercial-Sin obras derivadas 2.5 Argentina de Creative Commons. La autorización para su
publicación web fue cedida por el autor a la Facultad de Humanidades y Ciencias de la
Educación de la Universidad Nacional de La Plata. Este registro fue realizado por la
Biblioteca Prof. Guillermo Obiols de la Facultad de Humanidades y Ciencias de la Educación
de la Universidad Nacional de La Plata, Argentina." xml:lang="es" >
<META NAME="DC.type" CONTENT="bachelorThesis" xml:lang="en" >
<META NAME="DC.title" CONTENT="Metodología para la descripción de contenidos de imágenes
fotográficas" xml:lang="es" >
<META NAME="citation_abstract_html_url"
CONTENT="http://www.memoria.fahce.unlp.edu.ar/library?a=d&c=tesis&id=Jte378">
```

De modo que el tesoro, en cuanto SOC, tiene una función claramente definida: indizar y recuperar el conocimiento registrado. Esta función hace

² Registro de metadatos del repositorio Memoria Académica de la Facultad de Humanidades y Ciencias de la Educación de la Universidad Nacional de La Plata.

que el tesoro sea distinto, por ejemplo, de los bancos terminológicos, los glosarios y los diccionarios de sinónimos e ideas afines, que si bien muestran similitud, tienen otras finalidades.

El tesoro de segunda generación y sus nuevas funciones

La profesora canadiense Michèle Hudon recuerda que las primeras discusiones sobre las ventajas de la publicación electrónica de tesauros tuvieron lugar en la última década del siglo veinte y resaltaban principalmente la reducción de los costos, así como la expectativa de un incremento en la disponibilidad de tesauros en la Web y del uso de estos. Pero afirma Hudon: “Lo que era visto como una opción atractiva en 1995 se convirtió rápidamente en la única opción si es que los tesauros han de cumplir un rol en el nuevo mundo de la información” (Hudon, 2003: 116).

Hudon distingue entre los tesauros de primera generación y los tesauros de segunda generación. Los primeros son tesauros estáticos, en su mayoría impresos. Algunos han sido digitalizados y se encuentran en la Web como archivos PDF. Otros, si bien en formato html, no agregan mayores innovaciones ni facilitan su uso en diferentes aplicaciones.

Por el contrario, el tesoro de segunda generación debe ser dinámico, con registros almacenados en una base de datos accesible mediante los motores de búsqueda como Google o Yahoo. Los términos preferentes y las formas autorizadas del nombre deben estar hipervinculados entre sí mediante sus relaciones semánticas. El tesoro debe facilitar la navegación a través del orden sistemático y el orden alfabético. Tiene que ser fácil de consultar, intuitivo, con mínimas instrucciones y ayudas suficientes, pero simples. Es importantísimo que el tesoro pueda interactuar con distintas aplicaciones como las bases de datos bibliográficas, los sitios Web, etc. y estar disponible en los puestos de trabajo tanto del indizador como del usuario final.

Una de las funciones que cumple el tesoro en un sistema de recuperación moderno es su apoyo a la indización automática. Es una técnica basada en programas informáticos que seleccionan y procesan las palabras del texto de los documentos y las convierten en términos de indización. La indización automática se basa en el análisis lingüístico (morfológico, sintáctico y semántico) y en el análisis estadístico (frecuencia u ocurrencia de una palabra y su ponderación). El proceso es el siguiente (Cleveland & Cleveland, 2013; Gil Leiva, 2008):

- Eliminación de palabras vacías: el sistema descarta artículos, preposiciones, conjunciones, etc., mediante una lista de palabras vacías o *stopword* que la computadora pueda desechar.
- Análisis lingüístico: el sistema realiza la desambiguación de términos, la lematización (*steeming*) y el análisis tanto de la sintaxis como de la semántica.
- Análisis estadístico: cálculo de la frecuencia de palabras y frases en el texto y en la colección y ponderación de las mejores palabras y frases.

La función del tesoro en la indización automática consiste en facilitar la traducción de las palabras o frases seleccionadas por la computadora a los términos preferentes del tesoro. Esto se hace mediante métodos de mapeo que le dan consistencia a la indización, ya que la computadora selecciona palabras, pero no distingue conceptos.

Otra función del tesoro de segunda generación es la de facilitar la búsqueda expandida, que se define como la etapa del proceso de recuperación en la que se enriquece la pregunta inicial del usuario mediante la adición de términos que mejoran la recuperación. Almazán (2012) se refiere a esta función como *inducción*, dado que el tesoro guía o induce al usuario hacia otros términos que pueden interesarle para su búsqueda. Esta función de búsqueda expandida es útil porque los términos empleados por el usuario no siempre reflejan con exactitud su verdadero interés. El aporte del tesoro en este caso consiste en mapear los términos de búsqueda del usuario contra los términos preferentes del tesoro, sugiriendo las relaciones jerárquicas y asociativas correspondientes y facilitando la navegación.

En un estudio realizado por Shiri y Revie (2006) sobre la base de datos *CAB abstracts*, especializada en ciencias biológicas, se comprobó que el 60% de las búsquedas se expandieron mediante los términos del tesoro. De ellas, los usuarios seleccionaron principalmente relaciones jerárquicas y, en menor proporción, las asociativas. Sin embargo, los autores mencionados señalan que en otras investigaciones las relaciones asociativas han sido las preferidas por los usuarios para la expansión. Como quiera que sea, la búsqueda expandida por medio de un tesoro es una ventaja.

El tesoro puede cumplir además con otras funciones, como la verificación de la ortografía y la sugerencia de términos alternativos, como la típica

opción ¿Quiso decir ...? que se encuentra disponible en los motores de búsqueda de la Web o en los nuevos catálogos NextGen.

Hudon (2003) va más allá y propone que los tesauros deben considerarse imprescindibles no solo para las bases de datos bibliográficas disponibles en la Web, sino también para los portales temáticos e incluso para los motores de búsqueda de la Web.

Tipos de tesauros

Un *macrotesauro* es un tesoro de cobertura temática amplia, que suele tener un bajo nivel de especificidad, mientras que un *microtesauro* es un tesoro dedicado a un campo de conocimiento restringido, con un alto nivel de especificidad (Barité *et al.*, 2013). Un macrotesauro puede actuar como *paraguas* para interconectar varios microtesauros, de modo tal que los temas generales sean indizados por el macrotesauro y los específicos por alguno de los microtesauros.

Un *tesauro de búsqueda* es un tesoro que intenta asistir al usuario final durante la etapa de recuperación, aun cuando el sistema de información se maneje solo con lengua natural, para lo cual sugiere las posibles relaciones de equivalencia, jerarquía y asociación (Aitchison *et al.*, 2000; Bates, 1986, 1989). Un tesoro común, que se utiliza para la indización, puede usarse también como tesoro de búsqueda.

Un *tesauro multilingüe* es un tesoro en varios idiomas. El tesoro de UNBIS,³ creado por la Biblioteca Dag Hammarskjöld de la ONU, con sus versiones en árabe, chino, inglés, francés, ruso y español, es un ejemplo de tesoro multilingüe. Otros tesauros del sistema de la ONU cuentan también con esta característica.

Un *clasauro* es un SOC híbrido entre el tesoro y el sistema de clasificación. Introducido por Bhattacharyaa (1982), su finalidad es utilizar un solo SOC para la indización y la clasificación, es decir, para describir el contenido de los documentos en el catálogo o la bibliografía y para ordenar dichos documentos en el estante u otro tipo de ordenamiento sistemático. El *SOC Chinese Classified Thesaurus* de la Biblioteca Nacional de China es un ejemplo de clasauro.

³[http://lib-thesaurus.un.org/LIB/DHLUNBISThesaurus.nsf/\\$\\$searche](http://lib-thesaurus.un.org/LIB/DHLUNBISThesaurus.nsf/$$searche)

Autoevaluación

¿Qué es un tesaurus?

¿Cuáles son los componentes del tesaurus? Explique cada uno de ellos.

Describe las funciones del tesaurus en una base de datos bibliográfica según Foskett.

¿Qué función puede cumplir el tesaurus en la indización automática?

¿Qué función puede cumplir el tesaurus en la búsqueda expandida?

¿Qué tipos de tesauros podemos encontrar?

La historia del tesoro

En este capítulo trataremos sobre el origen del tesoro, que está íntimamente ligado a la aparición de la indización poscoordinada, y esta, a su vez, a la necesidad de procesar la enorme cantidad de información científica y tecnológica generada por la investigación en el siglo veinte.

El origen de la indización poscoordinada

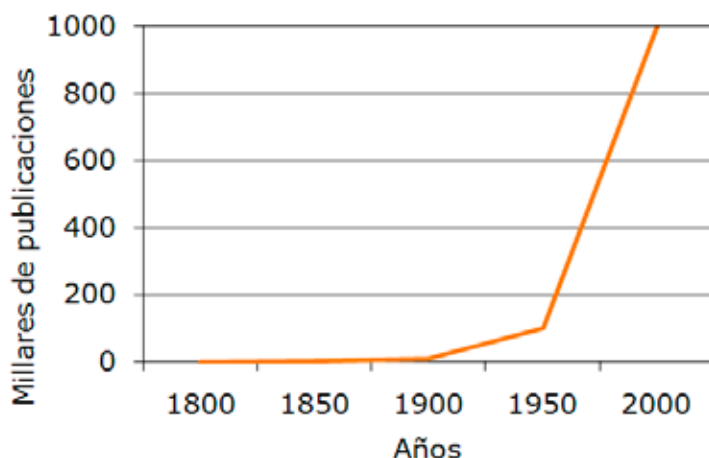
Si bien hubo algunos antecedentes previos, puede decirse que la indización poscoordinada comenzó hacia fines de la Segunda Guerra Mundial. Debido justamente a la guerra, se produjeron en los países beligerantes cantidades masivas de informes técnicos que registraban los resultados de investigaciones relacionadas con el conflicto y, hacia el final de la contienda, se sumaron a estos millones de documentos obtenidos en los países vencidos (Chu, 2003; Varlejs, 2004).

De igual modo, la posguerra trajo consigo una expansión aun mayor de la investigación científica y los desarrollos tecnológicos, estimulada por los altos presupuestos que destinaron a estos fines los gobiernos, las universidades y la industria. El resultado fue un nuevo flujo de conocimiento y, por extensión, una avalancha de publicaciones científicas. Este fenómeno es conocido como la *explosión de la información* y afectó principalmente a las revistas especializadas y a los sistemas de recuperación. A manera de ejemplo, entre 1910 y 1960 la literatura sobre economía creció de 5.000 a 40.000 artículos por año y la de psicología de 30.000 a 90.000 (ISSN Center, 2011; Riaz, 1989; van Dijk, 1972). Así, el número de publicaciones periódicas registró un crecimiento exponencial durante todo el siglo veinte (Figura 1).

Hacia 1970 las revistas científicas o académicas comunicaban al menos el 70% de los nuevos conocimientos y desarrollos tecnológicos (Mijailov & Guiliarevski, 1974). A esto había que agregar los informes técnicos, las

patentes, las normas, las tesis y un largo etcétera, que incluía además las imágenes fijas y en movimiento u otros materiales multimedia con contenido científico, por ejemplo, las fotografías de microorganismos tomadas a través de un microscopio, las imágenes del nacimiento de una estrella captadas por un telescopio, o el registro fílmico de un conflicto social.

Figura 1. Crecimiento de las publicaciones periódicas desde 1800 hasta 2000. Alrededor de 100 títulos en 1800, 1.000 títulos en 1850, d.10.000 en 1900, 100.000 en 1950 y 1.000.000 en 2000 (Fuentes: ISSN Center, 2011, Ríaz, 1989).



Antes de la Segunda Guerra Mundial, prácticamente la única forma de indización conocida era la asignación de epígrafes, pero llegó un momento en que esto no fue suficiente para el procesamiento de los documentos y se hizo necesario probar con algo nuevo. Las propuestas innovadoras culminaron en lo que hoy conocemos como indización poscoordinada (Kilgour, 1998).

Una de las primeras iniciativas en este sentido fue la del británico William E. Batten, quien se desempeñaba en el Britain's Imperial Chemical Industries, Patents and Intelligence Department. En 1944 Batten diseñó un sistema para indizar las patentes, basado en una ficha tabular cuya grilla o matriz de 20 filas por 20 columnas contaba con 400 celdas en total (Figura 2).

Las fichas diseñadas por Batten fueron conocidas como fichas *peek-a-boo* o fichas de coincidencia óptica.

En el sistema de Batten, a cada ficha se le asignaba un término que se extraía del propio documento. Entonces, si el documento n° 46 era la patente de un ventilador para autos, se tomaba la ficha del término *ventiladores* y se abría un orificio en el número de documento correspondiente (el 46 en la Figura 2). Luego se tomaba la ficha de *autos* y se hacía lo mismo. Cuando el usuario solicitaba información sobre ventiladores para autos, se tomaban las dos fichas, se superponía una a la otra y se contrastaban contra una fuente de luz. El rayo de luz solo podía pasar por los orificios que coincidían en ambas fichas, que era justamente el número de documento que incluía ambos conceptos, es decir, el 46.

Figura 2. Ejemplo de una ficha *peek-a-boo* de William E. Batten. Como se puede observar, el rayo de luz pasa por el orificio que se encuentra en la fila 40, columna 6, es decir, el documento 46, que trata sobre *ventiladores* y *autos*.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0				○																
20																				
40						○		○				○					○			
60	○								○									○		

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	○		○					○												
20								○						○						
40			○			○			○											
60																	○			

Otra propuesta fue la del estadounidense Calvin N. Mooers, quien en 1947 inventó una máquina a la que llamó Zator, capaz de coordinar dos o más fichas perforadas en el borde. Por cada documento ingresado en la colección, se abría una ficha Zatocard que soportaba un máximo de 40 perforaciones

en el borde (Figura 5.3). Cada perforación correspondía a un término, que Mooers denominó *descriptor*, de modo que se podían asignar hasta 40 descriptores por documento. El sistema Zatocode servía para una colección de 10.000 documentos o un poco más.

Para la recuperación, se marcaban en la máquina Zator¹ los descriptores que se deseaba recuperar. Todas las fichas se colocaban en un dispositivo de la máquina, semejante a una caja, que al accionarse comenzaba a moverse, agitando todas las fichas de modo tal que solo aquellas que respondían a los descriptores indicados caían fuera de la caja. El sistema producía algunas recuperaciones falsas, pero en general fue eficaz.

La tercera propuesta fue el sistema Uniterm del bibliotecario estadounidense Mortimer Taube. El sistema Uniterm era similar al de Batten, pero más sencillo. Alrededor de 1952, Taube diseñó una ficha con diez columnas numeradas del 0 al 9. Al igual que la ficha *peek-a-boo*, cada ficha Uniterm correspondía a un término (de ahí la palabra *uniterm* o unitérmino). Taube consideraba equivocadamente que un término era igual a una palabra, de manera que el concepto *escuelas públicas* daba lugar a dos fichas: una para *escuelas* y otra para *públicas*. Luego de varias experiencias y críticas, quedó claro que esta modalidad provocaba numerosas recuperaciones falsas, por lo que Taube debió modificar su criterio y aceptar que muchos términos son frases.

Por otro lado, cada documento ingresado en la colección recibía un número consecutivo que era anotado en la columna de la ficha, haciendo coincidir el último dígito del número de documento con el número de la columna. Por ejemplo, el documento 25 era anotado en la columna 5, el 282 en la columna 2, el 1318 en la columna 8 y así sucesivamente. Para la recuperación en el sistema Uniterm se tomaban las fichas de los términos que se deseaba coordinar y se verificaba qué números de documento coincidían. En la Figura 4 se muestra un ejemplo de la ficha Uniterm.

A partir del sistema Uniterm, Taube concibió la *teoría de la indización coordinada* sobre la cual dictó numerosas conferencias. El debate surgido alrededor de esta teoría llevó a la conclusión de que en todo sistema de recuperación es necesario coordinar los términos: en un caso, por ejemplo cuando

¹ Una fotografía de la máquina Zator puede verse en <http://www.flickr.com/photos/acrl/2387064291/>

Figura 4. Ejemplo de la fichas Uniterm de Mortimer Taube. El documento que trata sobre ventiladores y autos es el 46, que se encuentra en la columna 6.

AUTOS									
0	1	2	3	4	5	6	7	8	9
10	21	32	33	64	25	46	17	38	19
	51	82	43		125	86	37	98	49
	101		123		135	96	57		109
	131		233				87		

VENTILADORES									
0	1	2	3	4	5	6	7	8	9
20	31		53	84	35	46	27	48	29
	61		63	104	115	76	47		59
			133		165		67		
							107		

El origen del tesoro

Como solución, se presentaron en Estados Unidos tres propuestas que, si bien fueron formuladas por separado, coincidían en señalar la necesidad de crear un nuevo tipo de SOC, basado en términos bien definidos y ordenados sistemáticamente. Las tres propuestas también coincidieron en que ese SOC debía ser semejante al conocido diccionario de sinónimos e ideas afines *Roget's thesaurus of English words and phrases* y que debía acompañarse además de un índice alfabético (Aitchison & Dextre Clarke, 2004; Roberts, 1984).

Una de estas propuestas fue la de Calvin N. Mooers, a quien se acredita como autor del término *recuperación de información* y que, como ya vimos, fue el inventor del sistema Zatocode. En 1947 Mooers intercambió alguna correspondencia con varios colegas, proponiendo la creación de una lista de términos organizada sistemáticamente como el *Roget's thesaurus*. Denominó *descriptores* a los términos y *diccionario de descriptores* al nuevo SOC que proponía.

En el mismo año, Charles L. Bernier y Evan J. Crane del Chemical Abstract Service, sin conocer la propuesta de Mooers, hicieron un planteo similar y distinguieron tres tipos de relaciones: de subordinación, de coordinación y de equivalencia. Diez años después, el mismo Bernier sugirió la conexión de dos SOC de este tipo, en el sentido actual de macrotesauro y microtesauro.

Por último, en 1951 Hans P. Luhn, ingeniero de la IBM, propuso organizar los términos en familias conceptuales y agregar un índice alfabético. Aunque en algún momento Luhn se autoadjudicó la primicia del tesauro, se sabe que los otros autores mencionados le precedieron y que además él conocía la propuesta de Mooers. Tanto Luhn como Helen Brownson fueron los primeros en usar el término *tesauro* para referirse al nuevo SOC (Aitchison & Dextre Clarke, 2004; Roberts, 1984; Vickery, 1960b).

En 1957, B.E. Holm y L.E. Rasmussen, de la empresa E.I. du Pont de Nemours de Estados Unidos, iniciaron estudios para adoptar el sistema Uniterm de Mortimer Taube. No del todo satisfechos con este sistema y luego de varias modificaciones, en 1959 completaron el *Technical thesaurus of the E.I. du Pont de Nemours & Co. Engineering Department*, que, según Lancaster (2002), puede considerarse el primer tesauro que entró realmente en operaciones.

Las normas para el diseño y desarrollo de tesauros

Después del tesauro de du Pont se desarrollaron otros como el *Thesaurus of ASTIA descriptors* en 1960, el *Chemical engineering thesaurus* en 1961 y el *Thesaurus of engineering terms* en 1964, todos en Estados Unidos, que en general siguieron la metodología de du Pont.

Entre 1965 y 1967 se llevó a cabo el Proyecto LEX, coordinado por el U.S. Department of Defense, con la participación de otros organismos públicos y privados, en particular relacionados con la industria y la tecnología. El proyecto tuvo dos productos: por un lado el *Thesaurus of engineering and scientific terms* que fusionaba y ampliaba los tesauros mencionados en el párrafo anterior y por el otro, las primeras directrices para el diseño de tesauros, que fueron aprobadas por el Committee on Scientific and Technical Information (COSATI) y recomendadas para todas las agencias del gobierno estadounidense (Lancaster, 2002).

A partir de ahí comenzaron a desarrollarse diversas normas y directrices. Cabe aclarar que tanto las normas como las directrices son conjuntos de re-

glas e instrucciones que pueden seguirse para la ejecución de algo, en este caso del diseño y desarrollo de tesauros. La diferencia consiste en que una norma debe ser aprobada por un organismo normalizador como la International Organization of Standardization (ISO) en el ámbito internacional, o por los organismos nacionales correspondientes, como el Instituto Argentino de Normalización y Certificación (IRAM). En cambio, las directrices suelen ser recomendaciones de una institución de reconocido prestigio dentro de una especialidad, pero que no tiene el poder de aprobar normas, por ejemplo la Unesco o la IFLA.

Según Lancaster (2002), el desarrollo de las normas y directrices para el diseño de tesauros siguió dos corrientes: una estadounidense y otra británica. La primera se inició justamente con las directrices de COSATI en 1967. Sobre esta base, la Unesco redactó en 1971 sus directrices para tesauros monolingües, que tuvieron gran aceptación internacional (Unesco, 1971). Estas directrices influyeron en la redacción y aprobación de la primera norma internacional, que fue la ISO 2788 de 1974. Ese mismo año y también basada en COSATI, se redactó y aprobó la primera norma estadounidense, ANSI/NISO Z39.19:1974, que tuvo dos ediciones más en el siglo veinte y una cuarta en 2005.

Paralelamente, en 1969 se publicó en Gran Bretaña el *Thesaurofacet: a thesaurus and faceted classification for engineering and related subjects* (Aitchison *et al.*, 1969). Como su nombre lo indica, este tesoro se diseñó armonizando las directrices de COSATI con la metodología del análisis de facetas, creada por el bibliotecario indio Shiyali Ramamrita Ranganathan. El análisis de facetas tuvo desde el principio la fuerte adhesión de numerosos especialistas británicos que conformaron el CRG y para 1969 ya había sido aplicado al diseño tanto de sistemas de clasificación como a las listas de epígrafes (Coates, 1988).

Sobre la base de esta tradición británica, el organismo normalizador de ese país aprobó la norma BS 5723 de 1979; y dos miembros del CRG, Derek Austin y Peter Dale, prepararon la segunda edición de las directrices de Unesco, publicada en 1981 y traducida al español tres años después. La nueva edición de las directrices de la Unesco tuvo otra vez una marcada influencia en la revisión de la norma ISO 2788 y, consecuentemente, la segunda edición de esta última fue publicada en 1986 (ISO, 1986).

En 1985, la Unesco publicó las primeras directrices para tesauros multilingües, que dieron lugar a la norma ISO 5964 de ese mismo año (ISO, 1985). Más de dos décadas después, IFLA publicó sus directrices para este mismo tipo de tesauros (IFLA, 2009a).

No bien comenzó el siglo veintiuno y ante el avance de la tecnología informática, se inició la revisión de la norma estadounidense Z39.19, que se publicó en 2005 y fue revisada en 2010 (NISO, 2010). Por primera vez, una norma dejó de lado la presentación impresa y la gestión manual del tesoro para ocuparse principalmente de su presentación en línea, de las funciones que debe cumplir el *software* de gestión y de cómo debe orientarse hacia la interoperabilidad entre tesauros y otros SOC, incluyendo los tesauros multilingües, que dejan así de tener una norma separada. Además de este aporte innovador, la ANSI/NISO Z39.19:2005 amplió su cobertura hacia otros SOC como las listas de términos, los anillos de sinónimos y las taxonomías, que escapan al alcance de la presente obra.

Para no ser menos, el organismo normalizador británico, BSI Group, también decidió revisar su norma nacional. La nueva norma BS 8723-1/4 se publicó en cinco partes entre 2005 y 2008 (BSI Group, 2005-2008). Fiel a sus tradiciones, la norma británica conserva su orientación hacia el análisis de facetas, pero al igual que la estadounidense pone el énfasis en adecuar el tesoro a las nuevas tecnologías de la información y la comunicación. Además del tesoro, también esta norma incluye otros SOC como los sistemas de clasificación, las listas de epígrafes, las taxonomías y las ontologías, que no trataremos aquí.

Por último, la norma internacional también ha sido objeto de revisión. Las normas ISO 2788:1986 para tesauros monolingües y la ISO 5964:1985 para tesauros multilingües se dieron de baja y han sido sustituidas por una única norma publicada en dos partes: la ISO 25964-1/2. La parte 1, dedicada al tesoro, apareció en 2011 y la parte 2, dedicada a la interoperabilidad del tesoro con otros SOC, se publicó en 2013. Como sucedió anteriormente, la nueva norma ISO está basada en la norma británica (ISO 2011; 2013).

Los tesauros y las normas en Argentina

En nuestro país, a mediados de la década de 1970, el Centro Argentino de Información Científica y Técnica (CAICYT) reunió a varios especialistas

nacionales para conformar un grupo de trabajo encargado de elaborar un *Inventario de lenguajes documentarios no tradicionales* (CAICYT, 1980). El llamado Grupo CAICYT fue el principal promotor del Seminario Regional sobre Lenguajes de Indización, dictado por el profesor Frederick W. Lancaster (†), y auspiciado por la Unesco, que se llevó a cabo en Buenos Aires entre el 14 y el 26 de agosto de 1978 (Lancaster, 1985). Este seminario marcó un antes y un después respecto a los tesauros en nuestro medio, ya que el Grupo CAICYT se convirtió en un agente multiplicador de la metodología para el diseño y desarrollo de tesauros y organizó un nuevo seminario regional, destinado a docentes, que se realizó en 1982 (Molina & Zítara, 1998).

La consecuencia lógica de este impulso fue la aparición de los primeros tesauros argentinos, como el de recursos hídricos (Zítara, 1983) y el de psicología (García & Monfasani, 1984, 2009), considerados entre los mejores tesauros en lengua española (Álvaro Bermejo *et al.*, 1989a, b).

Un hecho trascendente, que situó a la Argentina entre los países que disponían de una norma nacional para tesauros monolingües, fue la publicación de la norma IRAM 32057:1983 (IRAM, 1983). Esta norma se basó principalmente en la primera edición de la ISO 2788:1974, debido a lo cual hoy en día se encuentra desactualizada (Martínez Tamayo & Mendes, 2010). Al cierre de esta edición, el Subcomité de Documentación de IRAM había iniciado las tareas correspondientes para adoptar una nueva norma IRAM-ISO 25964, que constituirá, sin duda, un valioso aporte al diseño y desarrollo de tesauros en nuestro país.

Autoevaluación

¿Qué situaciones contribuyeron al surgimiento de la indización poscoordinada?

Explique brevemente los sistemas de indización propuestos por Batten, Mooers y Taube.

Describa las propuestas que surgieron ante el problema de los unitérminos.

Explique cómo se inició el desarrollo de las normas y/o directrices para el diseño de tesauros.

¿Qué aportes hicieron las normas estadounidense, británica e internacional (ISO 25964) publicadas en el siglo veintiuno?

¿Qué suceso motivó la aparición de los primeros tesauros en Argentina?

¿Cuál es la situación de la norma argentina para el diseño y desarrollo de tesauros?

Diseño y desarrollo del tesaurus

En este capítulo trataremos sobre el diseño y desarrollo del tesaurus desde el punto de vista de la gestión de un proyecto. Para explicarlo nos permitimos recurrir al conocido Círculo de Deming y sus cuatro etapas: planificar, hacer, evaluar y actuar. Dentro de cada etapa, explicaremos algunas cuestiones esenciales para diseñar un tesaurus.

Consideraciones previas

Antes que nada, debe tenerse en cuenta que el diseño y desarrollo de un tesaurus suele ser muy costoso, tanto en dinero como en esfuerzo y en tiempo. Por lo tanto, antes de emprender un proyecto de esta envergadura, es necesario analizar las posibilidades de llevarlo a cabo. Un clásico análisis de fortalezas, debilidades, oportunidades y amenazas puede resultar adecuado para este fin. No obstante, se recomienda poner atención a las siguientes cuestiones (Aitchison *et al.*, 2000; ISO, 2011).

¿Existe un tesaurus sobre el dominio que interesa y en la lengua o lenguas que se necesita? Si la respuesta a esta pregunta es afirmativa, entonces conviene adquirir y utilizar ese tesaurus. Si la respuesta es negativa, se buscan otras soluciones.

¿Existe un tesaurus sobre el dominio, pero en otro idioma, que pueda traducirse? Si la respuesta es afirmativa, habrá que obtener los permisos correspondientes de los autores y editores antes de poner manos a la obra. De lo contrario, se continúa buscando otra solución.

¿Existe un tesaurus o varios que se puedan adaptar? La adaptación puede ser de distintos tipos: se puede derivar un tesaurus más específico (microtesaurus) de un tesaurus más general (macrotesaurus); también se pueden fusionar dos o más tesauros ya existentes que por separado no satisfacen todas las necesidades; convertir un sistema de clasificación en un tesaurus, o recurrir

a otros métodos de interoperabilidad (Martínez Tamayo *et al.*, 2011b). Si la respuesta a esta pregunta es afirmativa, entonces también se procede a obtener los permisos correspondientes antes de emprender las acciones.

¿Se cuenta con los recursos suficientes para completar el tesauro y su publicación? Si se inician las tareas de diseñar un tesauro y este queda inconcluso por la falta de recursos, el costo habrá sido inmenso, porque se perderán los recursos invertidos sin haber logrado el objetivo.

¿Se dispondrá de recursos para actualizarlo periódicamente en el futuro? El vocabulario de un tesauro no es estático; por el contrario, es necesario revisarlo en forma continua para descartar los términos que han dejado de usarse, modificar los términos que así lo requieran e incorporar términos nuevos.

Si las respuestas a las preguntas anteriores lo justifican, entonces solo entonces se puede pensar en diseñar un nuevo tesauro.

Planificar el tesauro

Una vez tomada esta decisión, se deben tener en cuenta otras consideraciones que hacen a la gestión del proyecto.

Equipo de trabajo

El diseño y desarrollo de un tesauro no puede ser obra de una sola persona, pues se trata de un proyecto complejo. Se necesita conformar un equipo de trabajo multidisciplinario, que puede requerir:

- a. Un director responsable de la conducción y coordinación del proyecto y la obtención de los recursos.
- b. Profesionales de la BDCI con experiencia en el diseño y desarrollo de tesauros.
- c. Futuros usuarios. Un tesauro suele tener dos tipos de usuarios: los indizadores responsables de la indización y los usuarios finales que actúan en la recuperación. Es muy conveniente que representantes de ambos grupos participen activamente en el equipo de trabajo.
- d. Expertos en el dominio del tesauro, con conocimiento específico sobre la terminología de ese dominio.
- e. Informáticos y diseñadores gráficos que aporten la tecnología necesaria para poner a punto el *software* de gestión del tesauro, diseñar una versión

amigable y atractiva en línea y asesorar sobre otros temas de su incumbencia.

f. Si se decide que el tesoro sea multilingüe, se requerirá también la participación de al menos un traductor por cada idioma extranjero.

g. Es conveniente contar con la asesoría de lingüistas y terminólogos.

h. Se deberá disponer además de personal auxiliar suficiente para el trabajo de secretaría y gestión administrativa.

Todo esto, desde luego, en la medida justa que se necesite.

Los miembros del equipo de trabajo mencionados en los incisos a-d deben reunir ciertas cualidades como una buena cultura general, una sólida formación académica, conocimientos de idiomas extranjeros particularmente el inglés, aptitud para trabajar en equipo y conocimiento del dominio y la terminología del tesoro.

Respecto al conocimiento sobre el dominio del tesoro, existen dos cuestiones a tener en cuenta. El propósito de un tesoro es facilitar la indización y la recuperación del conocimiento registrado, que son actividades propias del profesional de la BDCI, es decir, este profesional es el que sabe qué es un tesoro y para qué se usa. Por otro lado, para diseñar un sistema de conceptos propio de un dominio, debe tenerse conocimiento de ese dominio, de modo que es imprescindible contar con los especialistas del área. Lo ideal sería que los profesionales de la BDCI tengan experiencia en el uso y el diseño de tesoros y que además estén familiarizados con el dominio y la terminología del tesoro, y que los expertos en el dominio del tesoro estén familiarizados con el uso y el diseño de tesoros.

Plan de acción

Una vez conformado el equipo de trabajo, este procederá a desarrollar el plan de acción del proyecto, especificando los objetivos, metas, resultados esperados, los procesos que habrán de implementarse, los plazos de realización y los recursos necesarios para llevarlo a cabo.

El plan también debe especificar a qué usuarios estará dirigido, así como el nivel de conocimiento que estos usuarios tienen tanto del dominio como del uso del tesoro. Por ejemplo, un tesoro sobre la enfermedad de Chagas para médicos será diferente a un tesoro sobre el mismo dominio dirigido a la comunidad en general.

También se deben determinar algunas características del tesoro: si será publicado en forma impresa, en línea o ambas; si habrá de integrarse a un sistema de recuperación particular o a una lista de autoridades de nombres; si debe ser compatible con otros SOC; si se aplicarán estrictamente las recomendaciones de las normas o se harán excepciones, etc.

Los recursos

La provisión de recursos exige un compromiso institucional sólido y a largo plazo para que el proyecto sea sostenible y sustentable. La responsabilidad puede recaer en una sola institución o en varias, por ejemplo si el tesoro es parte de un proyecto colectivo. Los recursos a tener en cuenta son, entre otros, los siguientes:

a. Recursos financieros y físicos

Los recursos financieros deben asegurar tanto la finalización de una primera edición del tesoro, como su actualización periódica a lo largo de los años.

También se debe disponer de espacio físico adecuado, con el mobiliario y equipamiento tecnológico necesario y suficiente para que cada uno de los integrantes del equipo de trabajo pueda llevar a cabo su tarea.

b. Recursos bibliográficos

Es indispensable que el equipo de trabajo pueda contar con acceso permanente a las fuentes que permitan extraer y verificar la terminología del dominio y decidir otros aspectos, ya sean tratados, manuales, revistas especializadas, diccionarios, glosarios, enciclopedias, bases de datos bibliográficas, bancos terminológicos, nomenclaturas, etc. Los tesauros u otros SOC especializados en dominios relacionados también son de utilidad. Es preferible que estas fuentes estén disponibles en línea.

c. Normas y manuales

Para el diseño y desarrollo de tesauros conviene basarse en las normas que recomiendan las prácticas aceptadas y consensuadas. En nuestro país, en este momento, se hace necesario recomendar la norma internacional ISO 25964 (ISO, 2011), en la que nos basamos aquí. Como se indicó en el capítulo 5, actualmente el Subcomité de Documentación del IRAM se encuentra

realizando las tareas necesarias para adoptar como propia la mencionada norma, que será publicada como norma IRAM-ISO. Si bien las normas brindan una buena guía, se debe considerar que solo “pretenden recoger lo mejor de lo llevado a cabo hasta ahora, sin presuponer ni coartar los avances que se producirán en el futuro” (García Marco, 2006: 15). En otras palabras, la norma no se debe aplicar a ciegas, sino con una actitud crítica y teniendo en cuenta los avances posteriores.

Hay dos manuales clásicos que cuentan con varias ediciones: uno es el de los británicos Aitchison *et al.* (2000) y el otro es del estadounidense Lancaster (2002). Del ámbito iberoamericano se pueden mencionar el de Naumis Peña (2008), publicado en México, y los de Currás (1998) y Gil Urdiciain (2004) publicados en España. No obstante, debe tenerse en cuenta que ninguno de estos manuales ha tomado en cuenta las normas publicadas en el siglo veintiuno.

Todos los miembros del equipo multidisciplinario deben conocer en profundidad la norma y tener acceso permanente a la consulta de los manuales y otras fuentes.

d. Programa informático de gestión de tesauros

Para gestionar el tesoro es necesario crear una base de datos específica, que requiere un programa informático apropiado. Hasta hace unos años, las normas y manuales se referían a este recurso como una herramienta para producir tesauros impresos. Las normas publicadas en este siglo han puesto el énfasis en el tesoro electrónico, disponible en la Web. Para esto se recomienda un *software* que gestione la base de datos del tesoro, pero también que sea compatible con modelos de datos como SKOS o MADS y los programas de intercambio como el XML. En el capítulo 12 se tratan estos temas.

Hacer el tesoro

Los procesos de diseño y desarrollo del tesoro deben especificarse en la etapa de planificar y llevarse a cabo durante la etapa de hacer. Para implementar estos procesos es indispensable cumplir rigurosamente con el principio de garantía, que se explica en el Apartado 6.6. Hay experiencias de tesauros desarrollados por medios automatizados o semiautomatizados, aunque muchos requieren total o parcialmente el trabajo intelectual de los expertos.

Tradicionalmente, las normas y manuales han sugerido tres métodos para abordar el diseño y desarrollo de un tesoro. Como es costumbre en la BDCI, cada uno de ellos ha recibido variados nombres (AENOR, 1990; Aitchison *et al.*, 2000; BSI Group, 2005-2008; Gil Urdiciain, 2004; ISO, 2011; IRAM, 1983; Lancaster, 1985, 2002; Naumis Peña, 2008; NISO, 2010).

Método deductivo, teórico, sintético, a priori o top-down

Un grupo de expertos en el dominio del tesoro, basado en distintas obras de referencia especializadas, identifica aquellos conceptos más generales y a partir de ellos construye el orden sistemático de arriba hacia abajo, es decir, organizando los conceptos de lo general a lo específico. Una vez completado el orden sistemático, se procede a realizar el control del vocabulario de los términos y nombres propios incluidos.

Este método resulta útil para preparar el orden sistemático, pero presenta inconvenientes en los niveles jerárquicos más bajos, ya que al basarse en el juicio de los expertos, la colecta de términos puede resultar incompleta y las decisiones pueden apartarse del principio de garantía.

Método inductivo, empírico, analítico, a posteriori o bottom-up

Primero se colectan los términos y nombres propios de distintas fuentes bibliográficas, por alguna técnica de extracción de palabras claves o de estudios bibliométricos. Si se puede acceder a las historias de búsqueda de los usuarios finales de bases de datos relacionadas con el dominio del tesoro, también se extraen los términos de búsqueda. A medida que se van colectando los términos y nombres propios, se procede al control del vocabulario y se van agrupando para establecer los niveles jerárquicos de abajo hacia arriba, es decir, de lo específico a lo particular.

Gil Urdiciain (2004) recomienda la indización de unos 2000 documentos y la colecta de unos 15.000 términos antes de organizar las jerarquías. Aunque estas cifras dan idea del tamaño de un tesoro, debe tenerse en cuenta que hay tesauros con un número mayor o menor de términos, dependiendo de cada dominio. Lancaster (1985) relaciona la colecta con la disminución de términos nuevos. En la Figura 1 se muestra la curva propuesta por ese autor para vigilar el momento en que la cantidad de términos nuevos disminuye, que es el momento de comenzar a organizar las jerarquías.

Figura 1. Curva de crecimiento de términos. En el inicio de la colecta hay un crecimiento pronunciado porque todos los términos son nuevos, pero en cierto momento comienzan a repetirse y la incorporación de términos nuevos disminuye (Lancaster, 1985).



El método inductivo resulta efectivo para lograr un vocabulario completo y cumple estrictamente con el principio de garantía, pero al no contar con un orden sistemático previo, se complica la organización de los términos y nombres propios.

Método mixto o híbrido

Este método, fuertemente recomendado por la norma ISO 25964, combina los dos anteriores. Así, el grupo de expertos esboza un orden sistemático preliminar, mientras que la colecta de términos se realiza extrayendo los términos y nombres propios mediante procesos informáticos.

Los términos y nombres propios seleccionados son controlados y distribuidos en las distintas categorías ya establecidas en el orden sistemático preliminar. A medida que se avanza en la tarea, se van realizando los ajustes en el orden sistemático esbozado, de modo que todos los términos y nombres propios sean ordenados, aun cuando no hayan sido tomados en cuenta inicialmente.

Este método mixto permite lograr un vocabulario completo y organizar

el orden sistemático con un estricto cumplimiento del principio de garantía.

Los procesos del diseño y desarrollo de un tesoro

A continuación se presentan los principales procesos del diseño y desarrollo de tesauros. Algunos autores se refieren a estos procesos como etapas o fases (Gil Urdiciain, 2004; Naumis Peña, 2008; Shearer, 2004), pero en este manual, como se explicó en el capítulo 2, los consideramos procesos de organización del conocimiento.

a) Definición y análisis del dominio

Como ya se mencionó, una de las características del tesoro es su condición de SOC especializado en un dominio. En consecuencia, un primer proceso a llevar a cabo es definir y analizar el dominio del tesoro.

Un *dominio* es el ámbito de una actividad, una esfera específica de conocimiento, influencia o actividad. Puede ser una disciplina, arte o profesión, una especialidad de estas o un campo multidisciplinario.

Hay distintas recomendaciones para definir el dominio de un tesoro, que se mencionan en el capítulo 9.

b) Control del vocabulario

El proceso de control del vocabulario incluye varias operaciones, cuyo propósito es lograr un vocabulario normalizado y unívoco, es decir, controlado. Algunas de estas operaciones pueden realizarse en forma simultánea.

- **Colecta:** se realiza mediante el método inductivo ya descripto. Los términos y nombres propios colectados se registran en la base de datos del tesoro como *candidatos*, es decir, están propuestos, pero aún no han sido analizados y aprobados. La extracción automática de palabras claves es una técnica útil para la colecta. También hay experiencias en la aplicación de métodos bibliométricos para la colecta y selección de términos candidatos.
- **Selección:** no todos los términos y nombres propios candidatos deben pasar al tesoro. De hecho, se deberá analizar cada candidato y decidir si corresponde o no incluirlo en el tesoro. Este análisis significa estudiar en profundidad cada concepto general y el término que lo designa, así como cada concepto individual y el nombre propio que lo designa.

- Control de homónimos o desambiguación: cada término que se incorpora al tesauro debe representar un solo concepto general, es decir, no puede haber un término con dos significados distintos. De igual manera, un concepto individual debe estar representado por un solo nombre propio.
- Control de sinónimos: si un término candidato tiene uno o más términos sinónimos, deberá elegirse uno y solo uno como término preferente y establecer una relación de equivalencia con los otros términos. De igual modo, si existen dos o más formas variantes para un nombre propio, deberá elegirse una y solo una como forma autorizada del nombre y establecer una relación de equivalencia con las otras variantes.
- Normalización: se debe redactar una forma normalizada para cada término preferente y cada forma autorizada del nombre, siguiendo una norma pertinente. La norma ISO 25964 aporta las recomendaciones para las formas y accidentes gramaticales de los términos preferentes. Para normalizar los nombres propios conviene recurrir mejor a una norma de catalogación.

En el capítulo 7 se describen las recomendaciones para el vocabulario del tesauro y la relación de equivalencia.

c) Determinación de las relaciones jerárquicas y asociativas

Una vez que se completa el control del vocabulario, se procede a establecer las relaciones jerárquicas y asociativas. Aitchison *et al.* (2000) se refieren a estas como relaciones de micronivel. En ambos casos se trata de una relación entre dos conceptos y por lo tanto ambos elementos de la relación son términos preferentes, o bien uno es un término preferente y el otro una forma autorizada del nombre. Se detallan en el capítulo 8.

d) Construcción del orden sistemático

Este proceso consiste en organizar los términos preferentes y las formas autorizadas del nombre en un esquema lógico que evidencie el sistema de conceptos del dominio. Aitchison *et al.* (2000) se refieren a este orden como las relaciones de macronivel.

Como se mencionó anteriormente, conviene esbozar el orden sistemático con anterioridad a la colecta de términos y nombres propios para facilitar la distribución de estos últimos en el orden sistemático y ajustar este a medida

que se procede en el análisis, hasta consolidar el orden sistemático definitivo.

El análisis de facetas es la metodología recomendada para llevar a cabo el proceso de construcción del orden sistemático del tesaurus y se describe en el capítulo 9.

e) Publicación del tesaurus

Hoy en día la tendencia es publicar el tesaurus en formato digital y no impreso, como era común hasta no hace mucho tiempo. De hecho, las normas de diseño y desarrollo de tesaurus publicadas en este siglo veintiuno apenas hacen referencia a la versión impresa.

Las recomendaciones para la publicación del tesaurus se presentan en el capítulo 10.

Evaluar el tesaurus

La tercera etapa del diseño y desarrollo de un tesaurus es su evaluación. Hay dos evaluaciones a tener en cuenta:

Evaluación previa a la publicación: antes de publicar la primera edición de un tesaurus se recomienda hacer una evaluación, a modo de prueba piloto, que permita efectuar los ajustes necesarios antes de ofrecerlo al público para su uso.

Evaluación periódica: el tesaurus debe ser evaluado a intervalos regulares, para poder eliminar los posibles errores y determinar mejoras tanto en el vocabulario y las relaciones del mismo, como en el programa informático, especialmente en lo que hace a la interfaz usuario-sistema. En la evaluación se deben tener en cuenta las críticas y sugerencias de los usuarios del tesaurus.

En el capítulo 10 se hace referencia a las posibles formas de evaluación.

Actuar para mejorar el tesaurus

Es la cuarta etapa del diseño y desarrollo de un tesaurus. Una vez que se conocen los resultados de la evaluación, se deben hacer los ajustes necesarios para implementar las mejoras que surjan de la misma.

En el capítulo 10 también se hace referencia al mejoramiento continuo del tesaurus, especialmente a su actualización.

Principio de garantía

El diseño y desarrollo de un tesoro no surge de la nada. Debido a que el diseñador del tesoro se ve obligado a tomar decisiones importantes en cada uno de los procesos y operaciones de su diseño y desarrollo, resulta imprescindible que estas cuenten con un respaldo intelectual que garantice su validez. Hace más de un siglo que Wyndham Hulme postuló las bases de este principio, indispensable para desarrollar cualquier SOC (Barité, 2009, 2011a,b).

La profesora canadiense Clare Beghtol (1986) ha definido con exactitud el principio de garantía para un sistema de clasificación:

En general, la garantía de un sistema de clasificación puede pensarse como la autoridad que un clasificacionista invoca primero para justificar y subsecuentemente para verificar decisiones acerca de qué clases/conceptos incluir en el sistema, en qué orden las clases/conceptos deben aparecer en los esquemas, en qué unidades se dividirán las clases/conceptos, hasta dónde debe proceder la subdivisión, cuánto y dónde se dispone de síntesis, si los órdenes de mención son estáticos o variables y cuestiones similares (Beghtol, 1986: 110).

Adaptando esta definición al tesoro, podemos afirmar que la garantía es la autoridad que invoca el diseñador del tesoro primero para justificar y luego para verificar sus decisiones acerca de los conceptos, términos y nombres propios que se han de incluir en el tesoro, las relaciones entre ellos, en qué lugar del orden sistemático deben aparecer, en qué unidades se dividirán y hasta qué nivel de especificidad debe completarse el proceso de división. Recuérdese que la síntesis y el orden de mención al que hace referencia Beghtol no se aplican al tesoro, por ser este un SOC poscoordinado.

El principio de garantía ha evolucionado desde la primera aproximación de Hulme y hoy en día se ha especificado como señala Barité (2009, 2011a, b):

Principio de garantía literaria, bibliográfica o documental: los conceptos que formarán parte del tesoro deben extraerse de una fuente bibliográfica especializada, ya sean tratados, manuales, artículos de revistas, informes técnicos, patentes, obras de referencia, etc., cualquiera sea su soporte físico. Los bancos terminológicos o las nomenclaturas, en particular los que proceden de organismos especializados en el dominio del tesoro, son también valiosos.

Principio de garantía del usuario: el tesoro debe tener en cuenta la terminología del usuario final. En este sentido, es muy útil revisar las historias de búsqueda (*transaction logs*) realizadas por los usuarios en bases de datos bibliográficas afines al tesoro que se desea desarrollar. También es conveniente consultar a los usuarios finales a través de encuestas, entrevistas u otros medios y contar con un grupo de ellos como miembros del equipo de trabajo.

Principio de garantía organizacional: el tesoro debe reflejar la terminología del organismo u organismos para los que se diseña. Este principio resulta particularmente importante cuando el tesoro se destina al sistema de gestión de conocimiento de una empresa, una agencia de gobierno u otros organismos.

Principio de garantía cultural: el tesoro debe respetar la terminología y las particularidades culturales de la comunidad académica, profesional o social a la que estará dirigido. En este sentido es importante tomar en cuenta aquellos términos que pudieran resultar discriminatorios u ofensivos para los usuarios finales (Todaro & Martínez, 2006a,b).

Autoevaluación

¿Qué debemos tener en cuenta antes de comenzar a diseñar un tesaurus?

Mencione cómo debe estar conformado el equipo de trabajo ideal para diseñar y desarrollar el tesaurus.

Indique qué recursos son necesarios para un proyecto de diseño y desarrollo de un tesaurus.

¿Qué normas y manuales conviene tener en cuenta para el diseño y desarrollo del tesaurus?

Describa las principales características que debe tener el *software* de gestión de tesaurus.

Describa los métodos deductivo, inductivo y mixto para el diseño y desarrollo de tesaurus.

¿Cuáles son los procesos más importantes del diseño y desarrollo de un tesaurus?

¿En qué consiste el principio de garantía?

El vocabulario del tesoro

En este capítulo presentaremos el vocabulario del tesoro. Este suele dividirse en dos: el vocabulario de indización y el vocabulario de entrada (Figura 1). El primero incluye los términos preferentes y las formas autorizadas del nombre que pueden usarse en la indización; el segundo incluye los sinónimos, cuasi sinónimos y formas variantes del nombre, que no pueden usarse en la indización pero que remiten al término o al nombre aceptados.

Figura 1. Vocabularios de indización y de entrada. Solo los términos preferentes y las formas autorizadas del nombre pueden usarse para la indización.



El término preferente

Es el signo más importante del tesoro. Un término preferente es “un término usado para designar un concepto cuando se indiza” (ISO, 2011). Como se indicó en el capítulo 3, un término es la designación verbal de un

concepto general, de modo que cada término preferente representa un concepto general (y solo uno) y cada concepto general es designado por un término preferente (y solo uno). El término preferente también se denomina *descriptor* muy difundido en nuestro país pero las normas, en general, prefieren *término preferente* como usamos aquí (BSI Group, 2005-2008; ISO, 2011; NISO, 2010).

Cuando se diseña un tesauro se debe elegir un único término para cada concepto; de ahí que se denomine *término preferente*, pues ha sido preferido sobre los demás sinónimos que pudieran corresponder.

Sin pretender ser exhaustivos, a continuación se presentan algunos ejemplos de términos preferentes:

- Cosas y sus partes físicas:

- animales salvajes
 - automóviles de carrera
 - patas
 - radiadores
 - ríos

- Materiales:

- abrasivos
 - acero inoxidable
 - celulosa
 - fósforo
 - mármol de carrara

- Actividades o procesos:

- auditoría externa
 - cocción
 - diagnóstico
 - gestión de calidad
 - oxidación

- Eventos o hechos:

- aniversarios
 - apertura de cuenta
 - carnavales
 - festividades religiosas
 - huelgas

- Propiedades de personas, cosas, materiales o acciones:
 - personalidad
 - elasticidad
 - resistencia
 - fortaleza
 - opacidad
- Disciplinas, oficios o campos temáticos:
 - botánica
 - carpintería
 - ciencias de la tierra
 - historia
 - ingeniería biomédica
 - paleontología
- Unidades de medición:
 - kilogramo
 - megabyte
 - metro cúbico
 - minuto
 - milímetro
- Tipos de personas y organizaciones:
 - jóvenes
 - centros de investigación
 - organismos internacionales
 - personas adoptadas
 - pintores

Formas gramaticales de los términos preferentes

Para la redacción de un término preferente se recomiendan tres formas gramaticales: un sustantivo, una frase nominal adjetivada o una frase nominal preposicional.

Un sustantivo, sin artículo precedente, es la forma ideal para un término preferente, porque es la más fácil de recuperar y se considera un término preferente simple. Ejemplos:

- alegría
- alimentos

estética
máquinas
vuelos

Sin embargo, es imposible que el vocabulario de un tesoro se forme solamente con términos preferentes simples, ya que en muchos casos se requiere agregar al sustantivo un modificador, directo o indirecto. Se trata entonces de términos preferentes compuestos o multipalabra. Un término multipalabra puede ser de dos tipos: una frase nominal adjetivada es aquella cuyo núcleo es un sustantivo y su modificador un adjetivo (modificador directo). Se trata, entonces, de términos preferentes multipalabra. Ejemplo:

asentamientos marginales
discos compactos
medicina alternativa
pensamiento crítico
vuelos espaciales

Una frase nominal preposicional es aquella cuyo núcleo es un sustantivo y su modificador es una frase preposicional (modificador indirecto). En lo posible, deben evitarse y sustituirse por frases nominales adjetivadas. Por ejemplo *bibliotecas de hospitales* podría sustituirse por bibliotecas hospitalarias. Además, las frases nominales preposicionales son las más afectadas por las reglas para fraccionar términos multipalabra, según se explica en el Apartado 7.6 de este capítulo. Cuando no es posible sustituirlo, entonces puede redactarse un término preferente multipalabra como los siguientes:

agencia de seguridad
atención al cliente
bibliotecas para ciegos
patinaje sobre hielo
vuelos de cabotaje

En ocasiones es necesario incorporar al tesoro frases que no se ajustan a las tres formas gramaticales mencionadas, pero que están aceptadas y son de

uso común en un determinado dominio. Deben aceptarse sin cambios porque los usuarios finales del tesoro harán sus búsquedas por esas frases. Algunos ejemplos son los siguientes:

cromatografía líquida de alto desempeño
factor II de crecimiento similar a la insulina
medios de comunicación de masas
organizaciones no gubernamentales
procesamiento automático de datos

Los adjetivos, solos o como primera palabra del término preferente, no se permiten a menos que sean términos aceptados en el dominio del tesoro. Es común encontrarlos en idiomas como el inglés, pero son raros en lengua española. Por ejemplo, el Tesoro de UNBIS acepta como términos preferentes:

libre autodeterminación de los pueblos
libre comercio
libre elección de empleo

La norma estadounidense ANSI/NISO Z39.19:2005 propone la inclusión de algunos adjetivos con el fin de armar frases nominales adjetivadas (NISO, 2010). Por ejemplo:

Pequeños

NA Combínese con un sustantivo, por ejemplo *animales pequeños*, *automóviles pequeños*, etc.

Esta opción es una invitación a la incoherencia y a la pérdida del control de vocabulario, ya que distintos indizadores podrían aplicarla con diferentes criterios. Por este motivo, la norma ISO 25964 prefiere no recomendar su práctica (ISO, 2011). En nuestra lengua, además, habría que tener cuidado con la coincidencia necesaria entre el género y el número de sustantivos y adjetivos.

Los adverbios tampoco están permitidos como única o primera palabra del término preferente, a menos que sean de uso común en el dominio del tesoro. Por ejemplo, el Tesoro de UNBIS acepta los siguientes términos preferentes:

no alineación
no proliferación nuclear
no violencia

Los verbos, en cualquier conjugación, tampoco pueden usarse como términos preferentes. En su lugar deben emplearse sustantivos; por ejemplo: fabricación y no *fabricar*, construcción y no *construir*, curación y no *curar*.

La frase coordinada, que une dos sustantivos con la conjunción *y*, si bien está permitida en las listas de epígrafes por su carácter precoordinado, no tiene sentido en un tesoro. Por ejemplo, los epígrafes diccionarios y enciclopedias o computadoras y mujeres, que están aceptados en las listas de epígrafes LEMB y ARMARC, en un tesoro deberían expresarse con cuatro términos preferentes:

computadoras
diccionarios
enciclopedias
mujeres

La coordinación de estos términos se hará en el momento de la búsqueda mediante el uso del operador booleano AND; por ejemplo, diccionarios AND enciclopedias, computadora AND mujeres, computadoras AND diccionarios.

Otra frase no permitida es la frase invertida, como *golf*, *campos de* o bien *oración*, *libros de*. Hace décadas se usaba en algunos SOC, sobre todo en inglés, con el propósito de anteponer el sustantivo, pero no tiene sentido en español. La norma ISO 25964 recomienda que, en cualquier idioma, se redacte el término preferente en el orden de la lengua natural (ISO, 2011). Por lo tanto, los ejemplos anteriores serían: campos de golf, libros de oración.

Singular y plural

La norma ISO 25964 establece distintos criterios para la redacción del término preferente en singular o en plural, según se trate de conceptos concretos o abstractos (ISO, 2011).

Si es un concepto concreto y el núcleo del término preferente es un sus-

tantivo cuantificable —que puede responder a la pregunta *¿cuántos/as?*— debe expresarse en plural. Por ejemplo:

ampollas de vidrio
empresas industriales
fármacos
libros infantiles
mascotas

Si es un concepto concreto y el núcleo del término preferente es un sustantivo no cuantificable —que puede responder a la pregunta *¿cuánto/a?*— se redacta en singular. Por ejemplo:

aluminio
calor
carbón vegetal
colesterol
cloruro de sodio

En el caso de las partes del cuerpo, hay un tratamiento especial: si en un mismo organismo hay dos o más partes iguales, el término preferente se redacta en plural; si hay una sola, se redacta en singular. Por ejemplo:

arteria aorta
arterias
brazos
cerebro
glándulas suprarrenales

Generalmente, los conceptos abstractos se representan en singular, como se ejemplifica a continuación:

- El concepto es un fenómeno. Por ejemplo:
felicidad
personalidad
verano

- El concepto es una propiedad. Por ejemplo:
 - claridad
 - conductividad
 - eficiencia
- El concepto es un sistema de creencias o una ideología. Por ejemplo:
 - budismo
 - cristianismo
 - socialismo
- El concepto es una acción. Por ejemplo:
 - circulación sanguínea
 - decoración de interiores
 - fusión
- El concepto es una disciplina. Por ejemplo:
 - administración de empresas
 - ingeniería civil
 - psicología

No obstante, cuando el concepto abstracto posee varias especies, el término preferente se redacta en plural:

ciencias biológicas
 ciencias exactas
 conjuntos
 ideologías
 reacciones químicas

Hay casos en los que el singular y el plural tienen distintos significados, por ejemplo *pintura* como arte y *pinturas* como materiales con los que se pinta. Si se considera que la diferencia no está clara y puede prestarse a confusiones, deben diferenciarse con una nota de alcance como se indica más adelante en el apartado 7.7.3.

Forma autorizada del nombre

Un nombre propio es la designación verbal de un concepto individual, que a su vez representa un objeto único, como puede ser una persona, una

institución, un país determinado, una marca comercial, etc. Al igual que los términos preferentes, debe seleccionarse una (y solo una) forma autorizada del nombre para cada concepto individual.

Hace unos años, tanto las normas como los manuales se referían a esta forma autorizada del nombre con el término *identificador* (Aitchison *et al.*, 2000; IRAM, 1983; Lancaster, 2002), pero las nuevas normas publicadas en el siglo veintiuno no solo han eliminado este término, sino que han cambiado su significado, definiéndolo como un número o símbolo que identifica al término preferente (BSI Group, 2005-2008; ISO, 2011; NISO, 2010).

Otros autores prefieren referirse a los nombres propios como *descriptores onomásticos*, *descriptores geográficos*, *descriptores temporales*, etc. (Gil Urdiciain, 2004), pero las normas y directrices vigentes nunca han usado estos términos. La citada norma ISO 25964 los llama simplemente *nombres propios*, pero este término no distingue entre la forma autorizada y las formas variantes del nombre, de modo que hemos preferido utilizar estos dos últimos términos, según las recomendaciones de IFLA para el control de autoridades de nombres propios (IFLA, 2009b).

Las formas autorizadas del nombre pueden ser de distinto tipo y a continuación se muestran algunos ejemplos:

- Nombre propio de persona:
 - Juan de la Cruz, Santo, 1542-1591
 - Mafalda (personaje de historieta)
 - McCartney, Paul, 1942-
 - Bikila, Abebe, 1932-1973
 - Sarmiento, Domingo Faustino, 1811-1888
- Nombre propio de dinastía o familia:
 - Casa de Borbón
 - Casa de Windsor
 - Dinastía Ming, 1368-1644
 - Dinastía Ptolomeica, 323-30 a.C.
 - Rockefeller (familia)
- Nombre propio de ente corporativo:
 - Aeropuerto El Tehuelche (Puerto Madryn, Argentina)
 - Colombia. Ministerio de Comercio, Industria y Turismo
 - Estudiantes de La Plata (club de fútbol)

- Mura (restaurante : La Plata)
- Organización Panamericana de la Salud
- Nombre propio de un objeto o marca:
 - Friendship 7 (nave espacial)
 - Libertad (fragata)
 - Red Door (perfume)
 - Santa María (carabela)
 - Tango 01 (avión)
- Nombre propio de acontecimiento o reunión:
 - Congreso Nacional Bioquímico (11 : 2009 : Mar del Plata)
 - Festival de Cine de San Sebastián (1953 : San Sebastián)
 - Revolución de Mayo, 1810
 - Tratado de Versalles, 1919
 - Guerra Mundial, 1939-1945
- Nombre propio de título:
 - Don Segundo Sombra
 - Inmunología e inmunquímica
 - Quijote
 - Rocky (película cinematográfica)
 - Social science citation index
- Nombre propio de espacio (lugar o accidente geográfico):
 - La Rioja (Argentina : provincia)
 - La Rioja (España : provincia)
 - Metrópolis (ciudad de historieta)
 - Plaza Italia (La Plata, Argentina)
 - Popocatépetl (volcán, México)
- Nombre propio de tiempo:
 - Barroco, 1600-1750
 - Edad Contemporánea, 1453-1789
 - Gran Depresión, 1929-1934
 - Período Hispánico, 1536-1810
 - Quattrocento, 1400-1499

Si el nombre propio incluye un artículo antes del sustantivo, la forma autorizada del nombre también debe incluirlo. Por ejemplo:

El Salvador
La Habana (Cuba)
Los Angeles Lakers (club de básquetbol)

Un tesoro puede mostrar numerosos nombres propios, pero es muy difícil, si no imposible, que pueda incluir todos los que se necesitan a la hora de indizar en una biblioteca. Por ejemplo, el *Tesoro de la Unesco* posee la forma autorizada del nombre Argentina, pero no para Misiones o Río Negro. Posee el término preferente dramaturgo, pero carece de un nombre propio para Alejandro Casona o para Roberto Arlt.

Para solucionar este problema, el diseñador del tesoro puede proporcionar una instrucción en una nota de alcance, por ejemplo:

Argentina

NA Úsese para obras sobre el país en general. Para las obras sobre una provincia o localidad, indícese por el nombre propio normalizado de acuerdo con una norma de catalogación, por ejemplo *Santa Fe (Argentina : provincia)*.

dramaturgos

NA Úsese para obras sobre dramaturgos en general. Para obras sobre un dramaturgo en particular indícese por el nombre propio normalizado de acuerdo con una norma de catalogación. Por ejemplo *Miller, Arthur, 1915-2007*.

También se puede tomar la decisión de no incluir ningún nombre propio y aconsejar el uso de una lista de autoridades de nombres.

Abreviaturas y siglas

Las abreviaturas no están permitidas en el tesoro, porque afectan la recuperación. Se debe preferir familia a *flia.*, provincia a *prov.*, policía a *pcia.*

Las siglas están permitidas cuando son más conocidas que el nombre *in extenso*, por ejemplo: UNESCO en lugar de Organización de las Naciones Unidas para la Educación la Ciencia y la Cultura, SIDA en lugar de síndrome de inmunodeficiencia adquirida o láser en lugar de *light amplification by stimulated emission of radiation*.

Términos multipalabra

Como se ha mencionado, los términos preferentes pueden constar de una o varias palabras. La nueva norma internacional ISO 25964 usa el término *multi-word term* o término multipalabra en sustitución de *término compuesto* que utilizaban las normas anteriores, razón por la cual lo preferimos aquí (ISO, 2011).

Los términos multipalabra pueden ser de dos tipos: una frase nominal adjetivada (escuelas públicas) o una frase nominal preposicional (bibliotecas para ciegos), las cuales están formadas por un núcleo y un modificador, directo o indirecto:

- Núcleo: siempre es un sustantivo y suele indicar el género al que pertenece el término. En los ejemplos del párrafo anterior, los núcleos son *escuelas* y *bibliotecas*, respectivamente.
- Modificador directo: es un adjetivo y suele representar la especie. En el caso del primer ejemplo, el modificador directo es *públicas*.
- Modificador indirecto: es una frase preposicional. En el segundo ejemplo, el modificador indirecto es *para ciegos*.

En general, la norma y los manuales recomiendan fraccionar estos términos multipalabra en sus elementos más simples, excepto cuando pueda afectar la comprensión del indizador y del usuario final. Sin embargo, como se mencionó, no todos los conceptos pueden expresarse con una sola palabra. Esta contradicción entre lo conveniente (una sola palabra) y lo necesario (dos o más palabras) genera problemas que deben resolverse (Atchison *et al.*, 2000; ISO, 2011; Lancaster, 2002; Naumis Peña, 2008).

La principal ventaja del término preferente multipalabra es que mejora la tasa de precisión en el momento de la búsqueda, evitando las recuperaciones falsas. Por ejemplo, el término preferente control de calidad permite obtener mejores resultados que la coordinación de los términos preferentes *control + calidad*, ya que esta última opción podría recuperar un texto en el que se leyera, por ejemplo, “el *control* de la ortografía es indispensable para la *calidad* del texto”, que no tiene nada que ver con el concepto ‘control de calidad’.

Volviendo entonces a la norma ISO 25964, queda claro que siempre que sea posible, el término multipalabra debe descomponerse en términos más simples. Esta acción se conoce como *fraccionamiento de términos multipa-*

labra y puede ser de dos tipos: a) fraccionamiento semántico o b) fraccionamiento sintáctico. Pero si el fraccionamiento afecta la comprensión, entonces se mantiene el término multipalabra.

A continuación se presentan las recomendaciones de la norma ISO 25964 para mantener o fraccionar los términos multipalabra. Solo constituyen una guía, pero no son de aplicación obligatoria, pues cada uno de los casos que se presentan puede dar lugar a numerosas excepciones, dependiendo principalmente de la conveniencia para los usuarios finales.

Mantenimiento del término multipalabra

En los siguientes casos debe mantenerse el término multipalabra:

- Cuando el término está aceptado y es de uso común entre los usuarios finales, según el dominio del tesoro. Por ejemplo:

cromatografía capilar electrocinética micelar

procesamiento digital de señales

- Cuando el fraccionamiento puede conducir a una pérdida de significado o ambigüedad; por ejemplo: alimentos + animales podría recuperar información sobre animales que sirven de alimento o sobre alimentos para animales.

- Los términos que incluyen un nombre propio. Por ejemplo:

enfermedad de Chagas

teorema de Pitágoras

- Los términos cuyos modificadores han perdido su significado original.

Por ejemplo:

mesas redondas

sillas de tijera

- Los términos cuyo modificador sugiere una semejanza, a modo de metáfora o símil. Por ejemplo:

estructuras de árbol no puede fraccionarse en estructuras + árboles, porque la recuperación no sería pertinente.

- Los términos multipalabra que al separarse producen otros términos que no llegan a cubrir el mismo concepto. Por ejemplo:

establecimientos de salud

piletas de natación

- Los términos en los que el núcleo posee un significado diferente en

ausencia del modificador. Por ejemplo:

animales de peluche

flores de papel

Fraccionamiento semántico

Un término —simple o multipalabra— que representa un concepto A, puede fraccionarse en varios términos preferentes que, si bien por separado representan conceptos diferentes, una vez poscoordinados pueden significar lo mismo que el concepto A. Por ejemplo: el término *termómetros* puede fraccionarse en los términos preferentes temperatura + medición + instrumentos, que deberían asignarse juntos para significar *termómetros*. Igualmente, los términos *fisicoquímica* y *biotecnología* pueden fraccionarse, respectivamente, en los términos preferentes física + química y biología + tecnología.

Esta técnica puede producir una fuerte pérdida de consistencia en la indización y de precisión en la recuperación, de modo que la norma ISO 25964 no recomienda su aplicación.

Fraccionamiento sintáctico

Los términos multipalabra pueden fraccionarse sintácticamente en los siguientes casos:

- Cuando el núcleo (subrayado) se refiere a una propiedad o parte y el modificador representa el conjunto o poseedor de tal propiedad o parte, se fracciona:

_____ motores de avión

_____ lámparas de cristal

_____ vagones de ferrocarril

- Los términos preferentes después del fraccionamiento serían:

aviones

cristal

ferrocarriles

lámparas

motores

vagones

- Contrariamente, cuando el núcleo es el todo o poseedor y el modificador es la parte o propiedad, el término multipalabra se mantiene:

____ aviones comerciales
____ cristal cortado
____ ferrocarriles eléctricos

- Cuando el núcleo es una acción transitiva y el modificador es el objeto sobre el que recae la acción, se fracciona:

____ consumo de agua
____ laminación de acero
____ cultivo de cereales

Los términos preferentes resultantes después del fraccionamiento serían:

acero
agua
cereales
consumo
cultivo

laminación

- Contrariamente, cuando el núcleo es el objeto y el modificador es una acción que recae sobre aquel, se mantiene el término multipalabra:

____ acero laminado
____ agua potable
____ cereales procesados

- Cuando el núcleo es una acción intransitiva y el modificador es el objeto que realiza la acción, se fracciona:

____ comportamiento del consumidor
____ crecimiento del maíz
____ oxidación de metales

Los términos preferentes resultantes después del fraccionamiento serían:

comportamiento
consumidores
crecimiento
maíz
metales
oxidación

- Contrariamente, cuando el núcleo es el objeto y el modificador es una

acción en la que aquel ha estado involucrado, se mantiene el término multipalabra:

____ consumidores informados

____ maíz desgranado

____ metales oxidados

Como se puede observar, las frases con modificador indirecto son las más afectadas por el fraccionamiento sintáctico.

Por otro lado, la norma ISO 25964 se hace eco de lo sugerido por Lancaster (2002); recomienda fraccionar las frases con dos modificadores directos. Por ejemplo:

cámaras cinematográficas

cámaras submarinas

y no cámaras cinematográficas submarinas

estructuras metálicas

estructuras prefabricadas

y no estructuras metálicas prefabricadas

vuelos espaciales

vuelos tripulados

y no vuelos espaciales tripulados

Fraccionamiento de nombres propios

Los nombres propios nunca se fraccionan. Solo los términos preferentes y los términos no preferentes están sujetos a estas recomendaciones.

Desambiguación: calificadores y notas

Como ya se ha mencionado, los términos preferentes deben ser unívocos. Hay casos en que los términos resultan ambiguos y se requiere desambiguarlos. Existen dos recursos para hacerlo: los calificadores y las notas.

Homógrafos y calificadores

Los homónimos son términos que poseen un mismo significante pero

distinto significado. Pueden ser fonéticos (homófonos) o gráficos (homógrafos). Estos últimos son términos de idéntica grafía, pero que designan conceptos diferentes. Para distinguir uno de otro en el contexto de un tesaurus, se agrega al término preferente un término aclaratorio entre paréntesis, que se denomina *calificador*. Por ejemplo:

diafragma (músculo)
diafragma (dispositivo intrauterino)

Mercurio (deidad)
Mercurio (planeta)
mercurio (metal)

pinturas (materiales)
pinturas (obras de arte)

tanques (contenedores)
tanques (vehículos militares)

El calificador es parte del término preferente y por lo tanto no se pueden separar. El indizador deberá asignarlos tal cual aparecen en el tesaurus.

Por lo general el calificador se agrega solamente cuando los homógrafos aparecen en el mismo tesaurus. Tomando el ejemplo de *mercurio*, es difícil que en un mismo tesaurus se necesite distinguir entre la deidad y el metal. Sin embargo, hoy en día la interoperabilidad entre dos o más tesaurus es cada vez más común y algunos consideran que es conveniente agregar el calificador aunque el homógrafo no aparezca en el mismo tesaurus, en previsión de un posible proceso de interoperabilidad (ISO, 2011).

Las notas

La norma ISO 25964 establece tres tipos de notas, independientes unas de otras: la nota de alcance, la nota de definición y la nota histórica. Las normas anteriores incluían los tres tipos dentro de la nota de alcance, pero ha sido necesario separarlas para desarrollar el modelo de datos que recomienda la norma y que se describe en el capítulo 12 de la presente obra. En

este modelo, la nota de alcance se asocia con el concepto y las otras notas con el término.

a) Nota de alcance

Se trata de “una nota que define o clarifica los límites semánticos de un concepto” (ISO, 2011: 10).

En otros países de lengua española se utilizan también los sinónimos *nota aclaratoria*, *nota de aplicación*, *nota explicativa* y *nota de indización* (AENOR, 1990; Gil Urdiciain, 2004; Gil Leiva, 2008). Nosotros preferimos mantener *nota de alcance*, que es la forma utilizada por la norma argentina IRAM 32057:1983 y en otros países latinoamericanos, como Uruguay y México (Barité *et al.*, 2013; IRAM, 1983; Naumis Peña, 2008).

No todos los términos preferentes requieren una nota de alcance, solo aquellos que puedan resultar ambiguos. La nota se indica bajo el término preferente y su etiqueta es NA. Se reconocen varios tipos de nota de alcance (Aitchison *et al.*, 2000; ISO, 2011; Lancaster, 2002):

- Nota de inclusión o limitación positiva. Ejemplo:

educación artística

NA Úsese para las obras que tratan sobre la enseñanza del dibujo, la gráfica, la pintura y la escultura.

becas de estudio

NA Úsese para las becas a estudiantes de los ciclos primario y secundario.

- Nota de exclusión o limitación negativa. Ejemplo:

desarrollo

NA Excluye las obras que tratan sobre el desarrollo experimental o individual.

personas desplazadas

NA Excluye a los refugiados políticos. Para estos últimos úsese *refugiados*.

- Nota de instrucción. Ejemplo:

arquitectos

NA Úsese para las obras que tratan sobre los arquitectos en general. Para las obras que tratan sobre un arquitecto en particular, indícese por el nombre propio normalizado de acuerdo con una norma de catalogación, por ejemplo *Mies van der Rohe, Ludwig, 1886-1969*.

anemia

NA Úsese para obras muy generales. Prefiéranse en lo posible los términos específicos, por ejemplo *anemia hemolítica*.

La nota de alcance puede ser recíproca, cuando involucra a varios conceptos. Por ejemplo:

teatro

NA Úsese para las obras sobre el teatro como género literario. Las obras sobre el local destinado a la representación teatral se indizan por *teatros*.

teatros

NA Úsese para las obras sobre el local destinado a la representación teatral. Las obras sobre el teatro como género literario se indizan por *teatro*.

b. Nota de definición

Como su nombre lo indica, la nota de definición proporciona la definición de un término y es utilizada cuando el mismo puede interpretarse de modo ambiguo o equívoco.

La norma internacional ISO 25964 recomienda que, si es necesaria una nota de definición, esta sea independiente de las otras notas (ISO, 2011). Modifica así las disposiciones de las normas previas, en las cuales la definición debía incluirse como una nota de alcance. La etiqueta para la nota de definición es DEF. Por ejemplo:

educación comparada

DEF Estudio comparativo entre las teorías y prácticas educativas vigen-

tes en diferentes países.

informática

DEF Ciencia que estudia los medios físicos y lógicos para el tratamiento de la información.

Las normas y manuales para el diseño de un tesauro coinciden en que no todos los términos preferentes requieren una nota de definición (Aitchison *et al.*, 2000; ISO, 2011; Lancaster, 2002). No obstante, algunos autores sugieren que sería conveniente que todos los términos preferentes contaran con una definición, e incluso —como se mencionó en el capítulo 3— se ha recomendado la definición por intensión (género próximo y diferencia específica) como conveniente no solo para desambiguar el término preferente, sino también para facilitar la identificación de las relaciones jerárquicas (Aguado *et al.*, 1999; Barité, 2000a; Hudon, 1996; Vickery, 1960a). De hecho, hay tesauros que efectivamente definen todos sus términos preferentes.

c) Nota histórica

La nota histórica se incluye bajo un término preferente para indicar desde cuándo se utiliza el término y, en caso de modificación, relaciona el o los términos preferentes que designaban anteriormente el mismo concepto.

Para la norma ISO 25964, la etiqueta de la nota histórica es HN (*historical note*). Las normas española y argentina, así como otras normas latinoamericanas vigentes, no contemplan todavía este tipo de nota, de modo que no se cuenta con una etiqueta normalizada en nuestra lengua (AENOR, 1990; IRAM, 1983; Oficina Nacional de Normalización, 2000; ICONTEC, 1998). En 2013, el Subcomité de Documentación de IRAM aprobó la etiqueta NH (nota histórica) para la norma IRAM-ISO 25963 que se encuentra en curso y así la usamos aquí. Los siguientes son ejemplos de notas históricas:

Chequia

NA Las obras sobre el territorio de Chequia, referidas a los años comprendidos entre 1918 y 1992 deben indizarse por Checoslovaquia.

DEF Estado europeo independiente fundado el 1 de enero de 1993, a partir de la división de Checoslovaquia.

NH Término incorporado el 1993-01-10.

diabetes mellitus tipo 1

UP diabetes juvenil

UP diabetes mellitus insulino dependiente

NH Sustituye a los términos en desuso diabetes juvenil y diabetes mellitus insulino dependiente, de acuerdo con la recomendación del Expert Committee on the Diagnosis and Classification of Diabetes Mellitus de la American Diabetes Association. Término incorporado en 1997-09-01.

Bajo un término preferente o una forma autorizada del nombre pueden aparecer, simultáneamente, notas de definición, de alcance e históricas.

Relación de equivalencia

Como señalamos anteriormente, el vocabulario del tesoro está formado por un vocabulario de indización y un vocabulario de entrada. El primero está integrado por los términos preferentes y las formas autorizadas del nombre; el segundo, por los términos no preferentes y las formas variantes del nombre que se incluyen en el tesoro solo para remitir, respectivamente, a los términos preferentes y a las formas autorizadas del nombre.

Un término no preferente es un sinónimo o cuasi sinónimo del término preferente, que no se puede asignar en la indización pero que remite al término preferente de un tesoro. La forma variante del nombre es una forma del nombre propio que no se puede asignar en la indización, pero que remite a la forma autorizada del nombre en un tesoro.

La relación entre un término preferente y un término no preferente, o entre una forma autorizada del nombre y una forma variante del nombre es una relación de equivalencia. Algunos autores la designan como *relación de sinonimia* (Naumis Peña, 2008), pero debe tenerse en cuenta que los términos considerados equivalentes en un tesoro no siempre son sinónimos en sentido estricto. Por otro lado, *relación de equivalencia* se encuentra aceptada por la norma argentina y otras en español (AENOR, 1990; ICONTEC, 1998; IRAM, 1983; Oficina Nacional de Normalización, 2000).

En la relación de equivalencia se usan dos etiquetas: a) el término no preferente o la forma variante del nombre remiten al término preferente o a la

forma autorizada del nombre con la etiqueta USE, y b) el término preferente y la forma autorizada del nombre informan sobre los términos no preferentes y las formas variantes del nombre con la etiqueta UP, que significa *usado por*. Por ejemplo:

cacahuete
USE maní

maní
UP *cacahuete*

Confederación Helvética
USE Suiza

Suiza
UP *Confederación Helvética*

Las etiquetas anteriores pueden sustituirse por los símbolos ® en lugar de USE y = en lugar de UP. Ejemplo:

anacardos
® castañas de cajú

castañas de cajú
= *anacardos*

Godoy Alcayaga, Lucila
® Mistral, Gabriela, 1889-1957

Mistral, Gabriela, 1889-1957
= *Godoy Alcayaga, Lucila*

Cuando el diseñador de un tesauro se enfrenta con un conjunto de sinónimos o cuasi sinónimos, debe seleccionar uno de ellos como término preferente. Para esto deberá guiarse estrictamente por el principio de garantía que se explica en el capítulo 6. Sin ánimo de ser exhaustivos, a continuación describimos algunos criterios para orientar la selección:

- Origen lingüístico, por ejemplo: *políglota* **USE** multilingüe.
- Nombre popular o científico: *Solanum tuberosum* **USE** papa.
- Nombre comercial o nombre común: *biromes* **USE** bolígrafos.
- Siglas o nombre desarrollado: *PVC* **USE** policloruro de vinilo.
- Términos actuales u obsoletos: *diabetes juvenil* **USE** diabetes tipo 1.
- Variaciones ortográficas: *sicología* **USE** psicología.
- Variaciones culturales: *parqueos* **USE** estacionamientos.
- Términos prestados: *estudio de mercado* **USE** marketing.

Los cuasi sinónimos son términos que, si bien no son sinónimos en un sentido estricto, en el tesauro son tratados como tales. Por ejemplo:

trastornos mentales
USE enfermedades mentales

enfermedades mentales
UP *trastornos mentales*

También pueden ser extremos de un *continuum* como algunos antónimos. Debe decidirse si se incorporan ambos términos o si uno de ellos representa al otro. Por ejemplo:

frío
USE calor

calor
UP *frío*

desnutrición
USE nutrición

nutrición
UP *desnutrición*

Otra función de la relación de equivalencia es el llamado *reenvío ascendente*. En este caso, la relación se establece entre un concepto que no es necesario incorporar al tesaurus, y otro concepto más general que lo incluya. Por ejemplo:

ballenas
USE cetáceos

cetáceos
UP *ballenas*

ingenieros mecánicos
USE ingenieros

ingenieros
UP *ingenieros mecánicos*

Dentro de la relación de equivalencia, también se incluye el reenvío de un término multipalabra fraccionado a los términos preferentes que deben pos-coordinarse y se agrega un símbolo +. Por ejemplo:

motores de camión
USE motores + camión

camiones
UP+ *motores de camión*
motores
UP+ *motores de camión*

Autoevaluación

Explique la diferencia entre vocabulario de indización y vocabulario de entrada.

¿Qué es un término preferente?

Describa las formas gramaticales recomendadas para un término preferente.

Mencione en qué casos se debe redactar el término preferente en singular o en plural.

Si se decide no incorporar nombres propios al tesoro: ¿qué prácticas se pueden seguir?

¿En qué casos se pueden usar abreviaturas o siglas en un tesoro?

¿En qué casos se debe mantener un término multipalabra?

¿En qué casos se debe fraccionar un término multipalabra?

Describa los recursos para desambiguar un término.

Explique la relación de equivalencia.

¿Qué etiquetas y símbolos indican una relación de equivalencia?

¿En qué casos se recomienda establecer una relación de equivalencia?

Las relaciones jerárquicas y asociativas

En capítulos anteriores hemos mencionado que una de las características distintivas del tesauro es su red de relaciones explícitas. En este nos referiremos a las relaciones jerárquicas y asociativas que forman parte del micronivel señalado por Aitchison *et al.* (2000).

Las relaciones jerárquicas y asociativas constituyen una red de referencias cruzadas que permite, tanto al indizador como al usuario final, elegir el término preferente o la forma autorizada del nombre que más convenga para indizar o para recuperar información (Aitchison *et al.*, 2000; BSI Group, 2005-2008; ISO, 2011; Lancaster, 2002).

Todas estas relaciones son explícitas y recíprocas, de modo tal que si el término A está relacionado con B, el término B también debe estar relacionado con A.

Relación jerárquica

La relación jerárquica se establece entre dos conceptos. Uno representa el género o el todo y se denomina *término genérico*; la etiqueta que lo precede es TG. El otro representa la especie, la parte o un ejemplo del anterior y se denomina *término específico*; su etiqueta es TE. Si bien los dos elementos de la relación están designados por términos preferentes o formas autorizadas del nombre, se encuentran en un distinto nivel de jerarquía. Ejemplo:

reptiles
TE serpientes

serpientes
TG reptiles

La etiqueta TG puede sustituirse por el símbolo < y la de TE por >. Ejemplo:

automóviles deportivos
< automóviles

automóviles
> automóviles deportivos

Las normas y manuales citados permiten, de manera opcional, indicar en todos los términos preferentes el *término tope*, es decir, el concepto más genérico de una cadena de conceptos (Barité *et al.*, 2013). En un tesauro facetado, el término tope es la faceta de primer nivel, según se explica en el capítulo 9. En España y en Cuba el término tope se conoce como *término cabecera* (AENOR, 1990; Oficina Nacional de Normalización, 2000). La etiqueta del término tope es TT. En el siguiente ejemplo, los términos preferentes muebles de oficina, escritorios, escritorios gerenciales y escritorios secretariales comparten un mismo TT muebles:

muebles

TE muebles de oficina

TE muebles del hogar

muebles de oficina

TT muebles

TG muebles

TE escritorios

TE mesas

escritorios

TT muebles

TG muebles de oficina

TE escritorios gerenciales

TE escritorios secretariales

escritorios gerenciales

TT muebles

TG escritorios

escritorios secretariales

TT muebles

TG escritorios

Todos los términos preferentes y las formas autorizadas del nombre de-

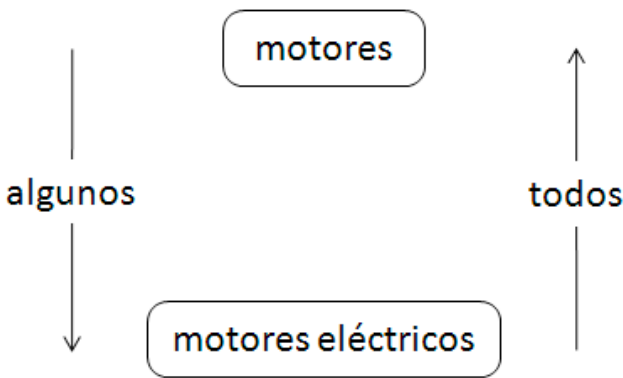
ben tener, al menos, una relación jerárquica. Si carecen de una relación jerárquica, se denominan *términos huérfanos*. Un tesauro no debe tener ningún término huérfano.

De acuerdo con la norma ISO 25964, hay cuatro tipos de relación jerárquica, que se describen a continuación (ISO, 2011).

Relación jerárquica género-especie

En este caso, el término genérico es el género y el término específico es la especie. También se conoce como *relación jerárquica genérica*. Es la más común de las relaciones jerárquicas y se puede verificar con la prueba que se muestra en la Figura 1.

Figura 1. Prueba de la relación jerárquica género-especie: Algunos motores son eléctricos, pero todos los motores eléctricos son motores. Por lo tanto, *motores* es el género y *motores eléctricos* la especie.



Algunos ejemplos de relación jerárquica género-especie son:

embarcaciones a vela
TG embarcaciones

embarcaciones
TE embarcaciones a vela

inmigrantes italianos
TG inmigrantes

inmigrantes
TE inmigrantes italianos

Aunque no es obligatorio, se puede agregar a la etiqueta de esta relación una

letra G para indicar que es una relación género-especie. En este caso sería TGG (término genérico género-especie) y TEG (término específico género-especie). Por ejemplo:

hospitales pediátricos	hospitales
TGG hospitales	TEG hospitales pediátricos

Relación jerárquica todo-parte

En esta relación, el término genérico representa el todo y el término específico representa una parte de ese todo. También se denomina *relación jerárquica partitiva*. La norma ISO 25964 solo permite cuatro tipos de relaciones todo-parte (ISO, 2011).

• Sistemas y órganos del cuerpo

ventrículos	corazón
TG corazón	TE ventrículos
pulmones	sistema respiratorio
TG sistema respiratorio	TE pulmones

• Lugares geográficos

Chubut	Argentina
TG Argentina	TE Chubut
Trelew	Chubut
TG Chubut	TE Trelew

• Disciplinas y sus especialidades

oncología	medicina
TG medicina	TE oncología
oncohematología	oncología
TG oncología	TE oncohematología

• Estructuras sociales

facultades
TG universidades

universidades
TE facultades

parroquias
TG diócesis

diócesis
TE parroquias

También aquí se puede agregar a la etiqueta de esta relación una letra P para indicar que es una relación todo-parte. En este caso sería TGP (término genérico partitivo) y TEP (término específico partitivo). Por ejemplo:

intestino delgado
TGP sistema digestivo

sistema digestivo
TEP intestino delgado

Facultad de Derecho
TGP Universidad de los Andes

Universidad de los Andes
TEP Facultad de Derecho

Relación jerárquica enumerativa o ejemplar

En esta relación, el término genérico es el género o el todo y el término específico es un ejemplo del anterior, designado por una forma autorizada del nombre. Por ejemplo:

Universidad de Palermo
TG universidades

universidades
TE Universidad de Palermo

Martí, José, 1853-1895
TG escritores

escritores
TE Martí, José, 1853-1895

La norma internacional ISO 25964 establece que la etiqueta opcional para la relación jerárquica enumerativa es BTI/NTI *broader term (instantial)*, *narrower term (instantial)*, pero las normas española, argentina y otras latinoamericanas vigentes no la han contemplado todavía (AENOR, 1990; Oficina Nacional de Normalización, 2000; ICONTEC, 1998; IRAM, 1983). En 2013 el Subcomité de Documentación del IRAM aprobó las etiquetas TGE/TEE (*término genérico enumerativo* y *término específico enumerativo*) para la norma IRAM-ISO 25963 que se encuentra en curso y así se usan en esta obra. Por ejemplo:

Lanín (volcán, Argentina)
TGE volcanes

volcanes
TEE Lanín (volcán, Argentina)

Atitlán (lago, Guatemala)
TGE lagos centroamericanos

lagos centroamericanos
TEE Atitlán (lago, Guatemala)

Relación polijerárquica

Es la relación que se establece entre un término preferente y dos o más términos genéricos, según el ejemplo que se muestra en la Figura 2.

Existen dos posiciones respecto a la relación polijerárquica. Algunos autores seguidores del análisis de facetas consideran que la estructura sistemática debe ser monojerárquica, es decir, cada concepto debe tener un solo concepto genérico. Aplicado al tesauro, significa que un término preferente solo puede tener un TG. Este enfoque coincide con el principio de *lugar de definición única* postulado por Farradane (1952) y revalorado más recientemente por otros autores (Gnoli, 2007).

Figura 2. Ejemplo de una relación polijerárquica. El término preferente órganos tiene en este caso dos términos genéricos: instrumentos de viento e instrumentos de teclado. Es, por tanto, una relación polijerárquica.



Contrariamente, en la estructura polijerárquica un concepto puede tener más de un concepto genérico, es decir, un término preferente podría mostrar más de un TG. Este enfoque está aceptado por las especificaciones de la norma ISO 25964, aunque se aclara que solo ocurre en algunos casos y debe tener un fundamento lógico (ISO, 2011). Una discusión sobre este tema se puede encontrar en el artículo de Miller (1997).

Relación asociativa

Es una asociación conceptual entre dos términos preferentes o formas autorizadas del nombre, que no se puede considerar como una relación de equivalencia ni como una relación jerárquica. Su reciprocidad es obligatoria y su etiqueta es TR (término relacionado). Por ejemplo:

aviadores
TR aviones

aviones
TR aviadores

La etiqueta TR puede sustituirse por el símbolo -. Por ejemplo:

acero laminado
- laminación

laminación
- acero laminado

La relación asociativa ofrece una gran libertad para relacionar los conceptos representados en el tesauro y su potencial ha sido valorado por diversos autores (Tudhope *et al.*, 2001).

Las normas y manuales publicados en el siglo pasado señalaban dos tipos de relación asociativa: la intracategoría y la intercategoría. Esta diferencia ha sido eliminada en la nueva norma internacional ISO 25964 que seguimos en el presente trabajo (ISO, 2011).

Esta norma internacional brinda, sin pretensión de ser exhaustiva, algunas orientaciones para establecer las relaciones asociativas.

Una disciplina o campo de estudio y su objeto o fenómeno estudiados.
Por ejemplo:

biología
TR seres vivos

seres vivos
TR biología

química
TR fenómenos químicos

fenómenos químicos
TR química

Una operación o proceso y su agente o instrumento. Por ejemplo:

administración
TR administradores

administradores
TR administración

procesamiento de datos
TR computadoras

computadoras
TR procesamiento de datos

Una acción y el producto de esa acción. Por ejemplo:

fotoduplicación
TR fotocopias

fotocopias
TR fotoduplicación

construcción
TR viviendas

viviendas
TR construcción

Una acción y su receptor u objetivo. Por ejemplo:

curación
TR enfermos

enfermos
TR curación

compra
TR proveedores

proveedores
TR compra

Un objeto o material y sus propiedades. Por ejemplo:

aluminio
TR maleabilidad

maleabilidad
TR aluminio

vino
TR color

color
TR vino

Un artefacto y sus partes, cuando estos no califican para una relación jerárquica todo-parte (véase el apartado 8.1.2). Por ejemplo:

bicicletas
TR manubrios

manubrios
TR bicicletas

gestión
TR planificación

planificación
TR gestión

Conceptos vinculados por una dependencia causal. Por ejemplo:

farmacodependencia
TR fármacos

fármacos
TR farmacodependencia

hábito de fumar
TR cigarrillos

cigarrillos
TR hábito de fumar

Un objeto o proceso y su contraagente. Por ejemplo:

plagas
TR plaguicidas

plaguicidas
TR plagas

infecciones
TR antibióticos

antibióticos
TR infecciones

Un objeto y su unidad de medida. Por ejemplo:

voltaje
TR voltio

voltio
TR voltaje

presión atmosférica
TR hectopascal

hectopascal
TR presión atmosférica

Un organismo o sustancia engendrado o derivado de otro. Por ejemplo:

mulas
TR burros

burros
TR mulas

chicle
TR Manilkara zapota

Manilkara zapota
TR chicle

Un término preferente puede tener todas o algunas de las relaciones semánticas aquí descritas. Por ejemplo:

papas

UP *patatas*

TT tubérculos

TG tubérculos comestibles

TE papas amarillas

papas coloradas

papas negras

papines

TR puré de papas

patatas

USE papas

Lo mismo sucede con las formas autorizadas del nombre. Por ejemplo:

Borges, Jorge Luis, 1899-1986

NA Comparte el sinónimo colectivo Honorio Bustos Domecq con el autor Adolfo Bioy Casares.

TT escritores

TG escritores argentinos

TR Bioy Casares, Adolfo, 1914-1999

Autoevaluación

Explique la relación jerárquica.

¿Qué etiquetas y símbolos indican una relación jerárquica?

¿Qué tipos de relación jerárquica reconoce la norma ISO 25964?

¿Qué es un término huérfano? ¿Se pueden aceptar términos huérfanos en un tesauro?

Explique la relación asociativa.

¿Qué etiquetas y símbolos indican una relación asociativa?

Mencione en qué casos se recomienda establecer una relación asociativa.

El orden sistemático del tesoro

Nos referiremos aquí a las relaciones de macronivel, es decir, al orden sistemático del tesoro (Aitchison *et al.*, 2000). Se trata de construir un esquema o plan de organización de los conceptos correspondientes al dominio del tesoro. Esos conceptos, designados por los términos preferentes y las formas autorizadas del nombre, se agrupan lógicamente de acuerdo con sus semejanzas y se separan por sus diferencias, de modo similar a un sistema de clasificación. Para lograr este orden sistemático también es indispensable cumplir con el principio de garantía, que se explicó en el Apartado 6.6.

Definición y análisis del dominio

Un *dominio* es una esfera o ámbito específico de conocimiento, influencia o actividad. Puede ser una disciplina, arte o profesión, una especialidad de estas o un campo multi o interdisciplinario. Una vez que se ha tomado la decisión de diseñar un tesoro, el primer paso es definir su dominio con exactitud.

Tennis (2003) propone dos ejes analíticos para definir un dominio:

- Eje 1 - Áreas de modulación: establece la extensión y el nombre del dominio. La extensión es el alcance total del dominio, es decir, su cobertura: lo que incluye y lo que no incluye. Por su parte el nombre designa al dominio, es la denominación por la cual debe llamarse ese dominio.
- Eje 2 - Grados de especialización: establece la focalización y la intersección del dominio. Mediante la focalización, la extensión del dominio disminuye, mientras que su intensidad aumenta. En otras palabras, mientras más restringido sea el alcance del dominio, más profunda será su especificidad. Por su parte, la intersección ubica un dominio frente a otros dominios. Puede ocurrir que un dominio esté claramente separado de otros, o bien que se interseque con alguno, dando lugar a un nuevo dominio con un nuevo nombre, como sucede con las interdisciplinas.

Una vez definido el dominio del tesoro, es necesario analizarlo. Hjørland y Albrechtsen (1995) han señalado que la mejor manera de estudiar la información o el conocimiento registrado en el área de la BDCI es analizar los dominios de conocimiento como comunidades de pensamiento o de discurso que son parte de la división social del trabajo. La organización del conocimiento, la estructura, los patrones de cooperación, las formas lingüísticas y de comunicación, los SAR y los criterios de relevancia son según los autores citados reflejos de los objetos de trabajo de estas comunidades y de su rol en la sociedad.

Barité *et al.* definen el análisis de dominio de la siguiente manera:

Denominación genérica de un conjunto de metodologías que tienen como objetivo común establecer las fronteras de un espacio especializado del saber, y determinar su estructura interna de conceptos y relaciones, con el propósito de representar el conocimiento que le es propio a través de una estructura conceptual o de otras modalidades. El análisis de dominio se realiza a partir de la identificación de los problemas y los tópicos propios del área, tal como se expresan en su literatura de referencia. Las aproximaciones pueden ser cuantitativas, cualitativas y mixtas. Estas metodologías se utilizan tanto para elaborar sistemas de organización del conocimiento y otras estructuras, como para revisarlas y evaluarlas (Barité *et al.*, 2013).

Entre los métodos aportados por la BDCI para el análisis de dominio, Hjørland & Albrechtsen (1995) mencionan el análisis de facetas que explicaremos aquí, los estudios bibliométricos, los estudios de usuarios particularmente de las conductas de búsqueda y algunas ideas de Bliss sobre la identificación de la estructura del dominio.

Por otro lado, en un tesoro hay algunos temas que no forman parte del dominio, pero que suelen ser necesarios. Vickery (1960a) los llama *temas marginales* y se debe decidir si se incluyen o no. Si se incluyen, es preciso considerar hasta qué nivel de especificidad se deben desarrollar, aunque generalmente se opta por un nivel menos específico que aquel del dominio. Algunos temas marginales comunes a casi todos los tesoros son los siguientes:

- Legislación
- Historia
- Educación
- Investigación
- Tecnología
- Administración
- Estadística
- Obras de referencia, publicaciones
- Bibliotecas, archivos, museos
- Sociedades, asociaciones

Análisis de facetas

En el capítulo 6 explicamos que el método inductivo es ideal para realizar la colecta de términos y nombres propios. Por otro lado, el método deductivo permite construir el orden sistemático para proceder a distribuir en ese esquema los términos y nombres propios colectados. Ambos métodos se complementan y la metodología recomendada para esto es justamente el análisis de facetas.

Entonces, el análisis de facetas consiste en organizar los conceptos, designados por los términos preferentes y formas autorizadas del nombre, en facetas homogéneas mutuamente excluyentes, derivadas de un universo común (Vickery, 1960a).

El creador del análisis de facetas fue el bibliotecario y profesor indio Shiyali Ramamrita Ranganathan, quien aplicó sus ideas en un sistema de clasificación de su autoría titulado *Colon classification* y expuso su teoría en el libro *Prolegomena to library classification* (Ranganathan, 1933, 1937). Desde mediados del siglo veinte, un grupo de expertos británicos —seguidores de Ranganathan y reunidos en el CRG— han revisado el método del análisis de facetas, aplicándolo a todo tipo de SOC (Broughton, 2011).

Aislados y focos

Un concepto que ha sido seleccionado para un SOC pero que aún no ha sido incluido en el esquema del orden sistemático, se conoce en el análisis de facetas como *aislado*. Un aislado es, entonces, “un concepto simple que puede situarse en varios contextos diferentes, conforme a sus características

intrínsecas y que es susceptible de ser incluido dentro de una faceta” (Barité *et al.*, 2013). Cuando el aislado se incorpora a una faceta, se convierte en un *foco*.

Si se aplica a un tesoro, el análisis de facetas consiste básicamente en convertir a los aislados en focos, es decir, en organizar los términos preferentes y las formas autorizadas del nombre en un esquema facetado.

Siguiendo a Vickery (1960a) y a Mills (1960, 2004), las etapas del análisis de facetas para un tesoro son:

- Identificación de las categorías apropiadas para el dominio
- División del dominio en facetas amplias
- División de cada faceta en subfacetas específicas
- Orden de clasificación de las facetas y subfacetas
- Adición de una notación
- Vinculación con la sección alfabética

Categorías y facetas

En el campo de la organización del conocimiento, Barité ofrece la siguiente definición de *categorías*:

Son abstracciones simplificadas, con fuerza de herramientas intelectuales, usadas por los clasificacionistas para indagar las regularidades de los objetos del mundo físico y del ideal y de las nociones [conceptos] que los representan. Este análisis y representación es realizado con miras a organizar lógicamente los sistemas de conceptos, suficientes para la organización del conocimiento en términos generales y el análisis de materia o clasificación de documentos en términos específicos (Barité, 2000b: 5).

Ranganathan (1967) estableció cinco categorías fundamentales que denominó *personalidad, materia, energía, espacio y tiempo* (PMEST). Para este autor, las cinco categorías, en ese orden estricto, podían ser aplicadas a todos los dominios. Por el contrario, el CRG postuló que las categorías a emplear deben surgir del análisis del propio dominio, con lo cual pueden o no coincidir de un dominio a otro (Mills, 1960, 2004; Spiteri, 1998; Vickery, 1960a). No obstante, se reconoce que ciertas categorías aparecen frecuentemente. Aitchison *et al.* (2000) las resumen como sigue:

a) Entidades, cosas, objetos

(por características)

Entidades abstractas, por ejemplo: ideas, disciplinas, fundamentos.

Entidades naturales, por ejemplo: radiación, nubes, ríos.

Entidades vivas, organismos; por ejemplo: bacterias, virus, mamíferos.

Artefactos, por ejemplo: manufacturas, obras de arte.

Atributos, propiedades, cualidades, estados, condiciones; por ejemplo: temperatura, color, confiabilidad, tamaño.

Sustancias materiales, sustancias componentes; por ejemplo: minerales, madera, cuero, plásticos, drogas.

Partes, componentes; por ejemplo: partes del cuerpo, techos, pisos.

Entidades completas, entidades complejas; por ejemplo: motores de vehículos, barcos, edificios, escuelas.

(por funciones)

Agentes (individuos, personal, organizaciones que realizan una acción), por ejemplo: artistas, marineros, médicos, asociaciones profesionales.

Agentes (equipamiento, aparatos que realizan una acción), por ejemplo: herramientas, computadoras, respiradores, lavavajillas.

Pacientes (receptores de una acción), por ejemplo: víctimas, beneficiarios, audiencias.

Productos finales, por ejemplo: alimentos, residuos, resultados de la investigación.

b) Acciones, actividades

Procesos o funciones, por ejemplo: desarrollo, enfermedades, pensamiento, caminata.

Operaciones, por ejemplo: enseñanza, curación, gerenciamiento.

c) Espacio

Espacio geopolítico, por ejemplo: países, provincias, ciudades.

Espacio geofísico, por ejemplo: ríos, montañas, islas.

d) Tiempo

Períodos de duración, por ejemplo: siglos, décadas, años.

Recurrimos ahora a un ejemplo que nos permita explicar mejor el análisis de facetas. Supongamos que el dominio del tesoro es bibliotecología y que hemos analizado las fuentes pertinentes para cumplir con el principio de garantía. También hemos identificado y definido algunos conceptos relevantes, que hemos designado con ciertos términos. Controlamos ese vocabulario inicial y obtuvimos los siguientes términos preferentes, que son nuestros aislados.

bibliotecarios

DEF Profesionales que ejercen la bibliotecología.

bibliotecas

DEF Instituciones que se caracterizan por desarrollar una colección de documentos, organizar esa colección y prestar servicios a una comunidad de usuarios. Cuentan para esto con el aporte de personal directivo, profesional y auxiliar.

catalogadores

DEF Bibliotecarios que se especializan en el proceso de catalogación.

directivos

DEF Miembros del personal que cumplen la función de dirigir la biblioteca.

empleados

DEF Miembros del personal que cumplen una función auxiliar en la biblioteca.

informáticos

DEF Profesionales que ejercen la informática.

licenciados

DEF Profesionales que poseen el grado académico de Licenciado.

magisters

DEF Profesionales que poseen el grado académico de Magister.

personal

DEF Conjunto de personas que cumplen alguna función en la biblioteca.

procesos

DEF Conjunto de actividades transformadoras que convierten una entrada (recursos) en una salida (producto, servicio).

profesionales

DEF Miembros del personal de la biblioteca cuya función requiere el ejercicio de una profesión universitaria.

recursos

DEF Conjunto de bienes materiales, financieros y tecnológicos con que cuenta la biblioteca para alcanzar sus objetivos y producir sus servicios.

referencistas

DEF Bibliotecarios que se especializan en el proceso de referencia.

servicios

DEF Productos de la biblioteca que dan respuesta a las necesidades de los usuarios y que son el resultado de los procesos.

usuarios

DEF Personas u organizaciones que demandan servicios de la biblioteca.

Ahora bien, al analizar estos conceptos en relación con las categorías notamos varias cosas:

El aislado bibliotecas corresponde a la categoría *[entidades complejas]*, ya que se trata de una institución que consta de numerosos elementos (personas, bienes, ideas etc.) que deben funcionar en conjunto para el logro de los objetivos.

Los aislados bibliotecarios, catalogadores, referencistas, directivos, empleados, informáticos, licenciados, magisters, personal y profesionales corresponden a las personas que cumplen funciones en la biblioteca, es decir, que llevan a cabo una acción. Pertenecen por eso a la categoría *[agentes]*.

Notamos también que catalogadores y referencistas comparten el atributo de ser bibliotecarios. Por su parte, bibliotecarios, informáticos, licenciados y magisters son profesionales. Además, profesionales, directivos y empleados son miembros del personal. A su vez, personal es el aislado más amplio o genérico dentro de la categoría *[agentes]*.

El aislado usuarios se refiere a las personas u organizaciones que reciben las acciones realizadas por los agentes. Pertenecen entonces a la categoría *[pacientes]*.

El aislado procesos es un conjunto de actividades y por ende pertenece a la categoría *[acciones]*.

El aislado recursos pertenece a la categoría *[materiales]*.

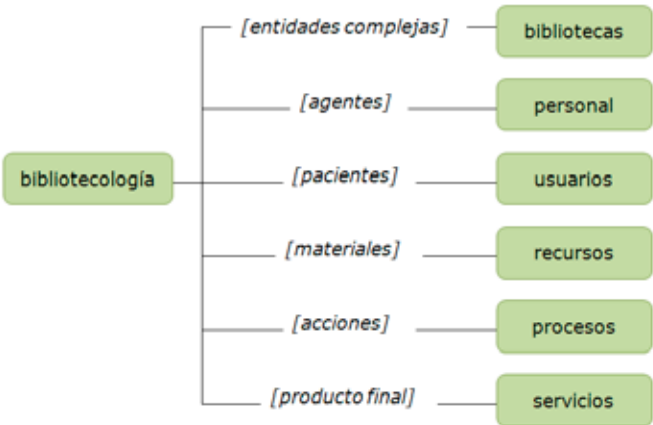
El aislado servicios pertenece a la categoría *[producto final]*.

Lo que hemos hecho es categorizar, dentro del dominio bibliotecología, un conjunto de aislados previamente seleccionados y controlados, que ya han sido aceptados con sus respectivos términos preferentes y definiciones.

El esquema resultante de esta primera etapa del análisis de facetas puede observarse en la Figura 1. Las facetas de primer nivel constituyen los términos tope de las relaciones jerárquicas del tesaurus.

Figura 1. Dominio, categorías y facetas. Esquema que muestra el dominio (bibliotecología), las categorías por las que se ha analizado ese dominio y las facetas resultantes, que constituyen el primer nivel del orden sistemático.

Nótese que las categorías se indican entre corchetes y con letra cursiva.

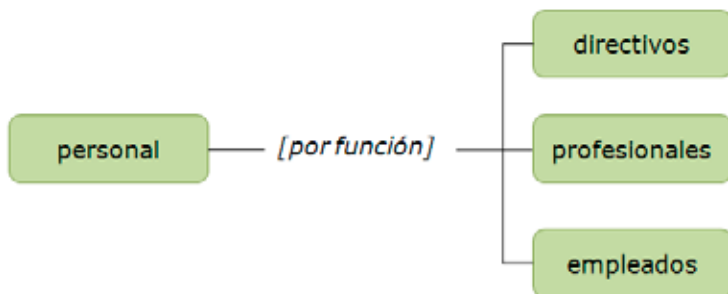


Una faceta se define como “el total de subclases que resulta de la aplicación de una única característica” (Mills, 1960: 8). O también: “conjunto total de subdivisiones de un concepto o clase, derivadas del mismo principio de división” (Barité *et al.*, 2013).

Subfacetas y focos

Seguimos ahora con los demás aislados de la categoría [*agentes*]. Encontramos que tres de ellos directivos, profesionales y empleados comparten el atributo de ser miembros del personal y por lo tanto podrían constituir una subfaceta de la faceta personal. Por otro lado, los tres conceptos se diferencian entre sí por la función que cumplen, y tomaremos este atributo como la característica de división [*por función*] (Figura 2).

Figura 2. Faceta, característica de división y subfacetas. Esquema que muestra una faceta (personal), una característica de división [*por función*] y una subfaceta de tres focos (directivos, profesionales, empleados). Nótese que la característica de división también se indica entre corchetes y con letra cursiva, pero a diferencia de la categoría, se le antepone la preposición *por*.



Podemos definir ahora tres elementos del análisis de facetas: atributo, característica de división y subfaceta.

Un *atributo* es una cualidad presente en un objeto. Algunos atributos permiten dividir un concepto y otros no. Ranganathan (1967) utilizó el siguiente ejemplo: la posesión del rostro es un atributo de los seres humanos, pero no permite la división porque todos poseemos un rostro, por lo

tanto es un atributo que no sirve para obtener conceptos más específicos. En cambio, el sexo, la edad, la estatura, el color del cabello son atributos que sí permiten la división en conceptos más específicos como varón y mujer, jóvenes y viejos, altos y bajos, rubios y morenos. El atributo de un concepto, elegido para obtener conceptos más específicos, es una *característica de división* (ISO, 2011; Ranganathan, 1967) o *principio de división* (Mills, 1960).

Una *subfaceta* se define como una “serie de subclases derivadas de una clase superior por la aplicación de una única característica” (Vickery, 1960a: 39). Se trata en definitiva del conjunto de focos que resultan de la aplicación de una única característica de división. Ranganathan llamó a esto *array*, mientras que el CRG prefirió *subfacet*. En español se ha traducido como *hileras*, *serie*, *rango* o *subfaceta*, que preferimos aquí (Barité *et al.*, 2013; Mills, 1960, 2004; Vickery, 1960a). Todos los focos de una subfaceta tienen el mismo nivel jerárquico, por eso la norma ISO 25964, refiriéndose específicamente al tesauro, denomina a los focos de una misma subfaceta *términos hermanos* y define *subfaceta* como el “conjunto de términos hermanos” (ISO, 2011: 1).

Continuamos analizando dos aislados más: bibliotecarios e informáticos. Ambos comparten el atributo de ser profesionales, pero se diferencian por la profesión que ejercen. Entonces, *[por profesión]* será la característica de división usada para ubicarlos en el esquema. Obtendremos así una subfaceta de tercer nivel (Figura 3).

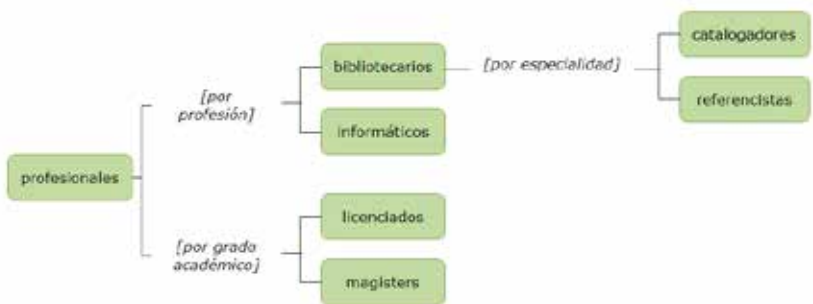
Sin embargo, también encontramos otros dos aislados licenciados y magisters cuyas definiciones claramente establecen que son profesionales, pero se distinguen uno de otro por el grado académico. Entonces tenemos otra característica de división, *[por grado académico]*, que aplicamos también a profesionales y obtenemos otra subfaceta de tercer nivel, distinta a la del párrafo anterior (Figura 3).

Es decir, se pueden aplicar una o más características de división a un mismo concepto, siempre y cuando cada una de ellas produzca una subfaceta con focos homogéneos y mutuamente excluyentes.

Tomemos ahora los aislados catalogadores y referencistas. De acuerdo con su definición, comparten el atributo de ser bibliotecarios y se diferencian por su especialidad. La característica de división será *[por especiali-*

dad]. El resultado es una subfaceta de cuarto nivel, que también se muestra en la Figura 3.

Figura 3. Subfacetas de segundo, tercero y cuarto nivel. Los aislados referidos a los distintos profesionales han sido organizados por dos características de división: *[por profesión]* y *[por grado académico]*, que resultan en dos subfacetas de tercer nivel. A su vez, los aislados referidos a los distintos tipos de bibliotecarios han sido organizados según la característica de división *[por especialidad]* para obtener una subfaceta de cuarto nivel.



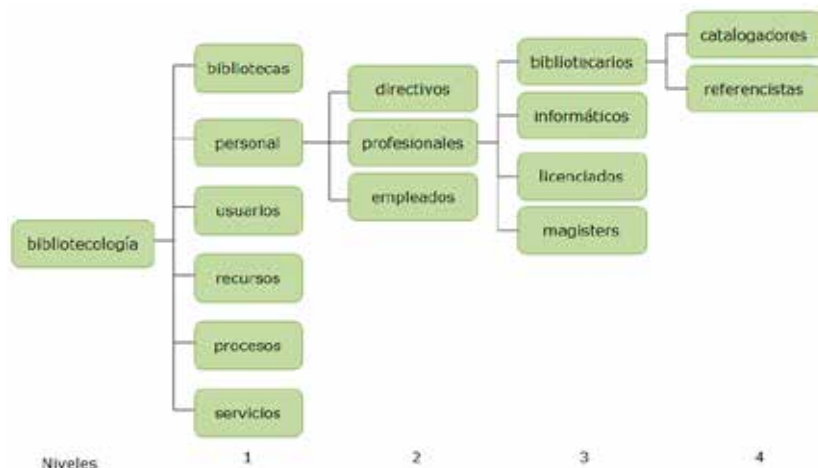
Una subfaceta debe incluir todos los focos que derivan de la característica de división. Para agotar la división, en un sistema de clasificación generalmente se utiliza una clase residual *otros*. Pero en un tesauro no puede haber un término preferente *otros*. Algunas veces, como afirma Mills (2004), se recurre a conceptos expresados por una frase negativa. Por ejemplo, si al término preferente *bebidas* se le aplica la característica de división *[por contenido de alcohol]* se obtendrán los términos preferentes *bebidas alcohólicas* y *bebidas no alcohólicas*. Este último concepto estaría incluyendo todos los tipos de bebidas que no son alcohólicas y con esto el nivel de exhaustividad de la subfaceta estaría completo.

Se procede igual con todos los aislados, es decir, con todos los conceptos designados por términos preferentes o formas autorizadas del nombre.

La cadena y las relaciones jerárquicas

Observemos ahora la Figura 4, donde se muestran todos los niveles del esquema referido al ejemplo.

Figura 4. Cadena. Esquema que muestra el dominio y las facetas y subfacetas, hasta el cuarto nivel. Se distinguen dos cadenas completas: a) personal, profesionales, bibliotecarios, catalogadores y 2) personal, profesionales, bibliotecarios y referencistas.



Los términos preferentes personal, profesionales, bibliotecarios y catalogadores forman una *cadena*. Los conceptos personal, profesionales, bibliotecarios y referencistas forman otra *cadena*. Entonces, podemos decir que una cadena es un conjunto de focos subordinados, derivados de la aplicación de distintas características de división. De acuerdo con Barité *et al.*, una cadena es:

Serie jerárquica de términos de extensión decreciente e intensión creciente derivada por subdivisiones sucesivas, producto de la aplicación de características cada vez más específicas. El término que se ubica en el extremo cenital de una cadena se llama *summum genus* y corresponde al término tope de un tesoro o a la clase de un sistema de clasificación. Lo que caracteriza a una cadena es que todos sus términos sin excepción, están en relación jerárquica unos respecto de otros (Barité *et al.*, 2013).

En la sección sistemática de un tesoro, las cadenas de la Figura 9.4 se indicarían como sigue (téngase en cuenta que solo se ha desarrollado la categoría *[agentes]*):

[agentes]

personal

[por función]

directivos

profesionales

[por profesión]

bibliotecarios

[por especialidad]

catalogadores

referencistas

informáticos

[por grado académico]

licenciados

magisters

empleados

[pacientes]

usuarios

[materiales]

recursos

[acciones]

procesos

[producto final]

servicios

En una cadena no se pueden saltar los conceptos de nivel intermedio. Por ejemplo:

Cadena incorrecta:

sistema circulatorio

aurículas

ventrículos

Cadena correcta:

sistema circulatorio

corazón

aurículas

ventrículos

Etiqueta de nodo

Como se puede observar en las figuras 9.1 a 9.4, se han indicado entre corchetes tanto las categorías como las características de división. La norma ISO 25964 llama a esta indicación *node label*, que se ha traducido al español como *indicador clasificador* (AENOR, 1990) y también como *etiqueta de nodo* (ICONTEC, 1998). El Subcomité de Documentación del IRAM adoptó *etiqueta de nodo* para la norma IRAM-ISO 26954-1, cuya preparación se encuentra en curso.

El término *etiqueta de nodo* se define como “una etiqueta que se inserta en el orden sistemático para mostrar cómo han sido ordenados los términos” (ISO, 2011: 7).

La etiqueta de nodo de la categoría se escribe entre corchetes y con letra cursiva, mientras que la etiqueta de nodo de la característica de división se escribe entre corchetes, con letra cursiva y anteponiendo la preposición *por*, como se observa en las figuras 9.1 a 9.4.

El orden de archivo u orden de intercalación

La metodología del análisis de facetas fue creada inicialmente para los sistemas de clasificación, y en ese caso deben completarse dos etapas más: a) la definición del orden de cita u orden de mención y b) la definición del orden de intercalación.

En un tesaurus no se requiere un orden de mención porque es un SOC poscoordinado. Sin embargo, se debe prestar atención al orden de intercalación. De acuerdo con Mills (1960, 2004) el orden de intercalación tiene dos componentes:

- Orden de las facetas y las subfacetas: es la sucesión en que las facetas y las subfacetas (cada una de ellas considerada como un bloque) se ordenan una tras otra. Este orden se basa en el principio de extensión decreciente, o, dicho de otra forma, de lo general a lo particular.
- Orden de los focos en una subfaceta: se basa en el principio de secuencia útil de Ranganathan, que se explica en el Apartado 9.3.

Principios del análisis de facetas

Como se mencionó anteriormente, la metodología del análisis de facetas procede de dos fuentes: Ranganathan y el CRG. Ambos han aportado numerosos principios que deben observarse cuando se intenta aplicar esta metodología al desarrollo de un SOC.

Ranganathan postuló 46 cánones, 13 postulados y 22 principios que se exponen en su libro *Prolegomena to library classification*. Por un lado, algunos de estos cánones, postulados y principios resultan reiterativos; por el otro, el estilo de redacción de Ranganathan no es fácil de leer.

El CRG, si bien se basa en los principios de Ranganathan, tiene sus propios aportes, particularmente en lo que se refiere a las categorías fundamentales. Desafortunadamente, el CRG nunca publicó un texto similar a *Prolegomena*, que reuniera la teoría del análisis de facetas según el enfoque de este grupo. Por el contrario, los principios se encuentran dispersos en numerosos artículos y otras fuentes.

Por lo antes expuesto, la profesora canadiense Louise Spiteri propuso compatibilizar los principios de Ranganathan con los del CRG, para simplificar su explicación y comprensión. A continuación se describen los principios del análisis de facetas de acuerdo con el modelo simplificado de Spiteri (1998).

Principios de la característica de división:

- Principio de diferenciación: una característica de división debe generar al menos dos focos mutuamente excluyentes.
- Principio de relevancia: una característica de división debe ser relevante para el propósito del tesoro.
- Principio de averiguabilidad: una característica de división se debe poder averiguar y definir.
- Principio de permanencia: una característica de división debe continuar sin cambios, mientras no se modifique el esquema.
- Principio de homogeneidad y exclusión mutua: una característica de división debe producir focos homogéneos y mutuamente excluyentes, es decir, los contenidos de un foco no pueden superponerse con los de ningún otro y cada foco debe proceder de una única característica de división.
- Principio de las categorías fundamentales: no existen categorías fundamentales para todos los dominios. Las categorías deben basarse en la naturaleza del dominio a ser analizado.

Principios del orden de las facetas, subfacetas y focos:

- Principio de sucesión relevante: la sucesión de características de división debe ser relevante para el propósito del tesoro.
- Principio de sucesión consistente: la sucesión de características de división en el esquema debe mantenerse consistentemente, mientras el esquema no sea modificado.

Principios de la terminología:

- Principio del contexto: la denotación de un término debe estar determinada por las diferentes facetas, subfacetas y focos pertenecientes a la misma cadena del término en cuestión.
- Principio de actualidad: el término usado para designar un foco debe ser de uso corriente entre los especialistas del dominio del tesoro.

Principios de la notación:

- Principio de sinonimia: la notación debe ser única; no puede haber dos notaciones diferentes para un mismo concepto.

- Principio de homonimia: el concepto representado por una notación debe ser único; no puede haber dos conceptos representados por la misma notación.
- Principio de hospitalidad: el sistema notacional debe permitir la adición de nuevas facetas y focos en cualquier punto del esquema.

Principios del orden de clasificación:

- Los focos en la cadena deben mostrar una extensión decreciente, de lo general a lo particular.
- Los focos en la subfaceta deben ordenarse de acuerdo con alguno de los criterios de la secuencia útil expresados por Ranganathan, entre ellos: complejidad creciente o decreciente, concreción creciente o decreciente, cantidad creciente o decreciente, contigüidad espacial (lejanía o cercanía en el espacio), orden cronológico (lejanía o cercanía en el tiempo), orden evolutivo (lejanía o cercanía del origen), orden canónico, etc. De no existir otro criterio más conveniente, se puede recurrir al orden alfabético.

Orden sistemático por campos temáticos

La norma ISO 25964 acepta la construcción de un orden sistemático por campos temáticos en lugar de facetas. Así, un dominio se ordena de acuerdo con sus distintas especialidades u otras subáreas aceptadas en ese dominio. Por ejemplo, medicina podría dividirse en especialidades como medicina generalista, pediatría, gerontología, y sus distintas especialidades ginecología, inmunología, etc., o por subáreas como anatomía, fisiología, patología.

La norma también acepta un orden sistemático basado en campos temáticos que a su vez son subdivididos por facetas.

Notación

La notación es un conjunto de símbolos que representan un concepto en algunos SOC, especialmente en los sistemas de clasificación (ISO, 2011). Los símbolos pueden ser letras, números, signos de puntuación o una combinación de todos estos.

En un tesoro, la notación puede servir para dos cosas:

- Mecanización: dado que la notación tiene un valor ordinal, facilita la

organización de los conceptos del orden sistemático por medio de un programa informático.

- **Nexo:** en los tesauros impresos la notación sirve de nexo entre la sección alfabética y la sección sistemática. En los tesauros en línea esta función no es necesaria, dado que todos los términos preferentes y las formas autorizadas del nombre deben estar hipervinculados, de tal manera que un simple clic del *mouse* remita del orden alfabético al sistemático.

La notación debe reunir ciertas cualidades, que describe Foskett (1996):

- **Unicidad:** cada notación debe representar un concepto y solo uno, evitando la sinonimia y la homonimia.
- **Simplicidad:** la notación debe ser fácil de comprender, usar y recordar.
- **Brevedad:** la notación debe ser breve. Para ello es necesario que la notación de base, es decir, la notación que corresponde a las facetas de primer nivel, sea lo más corta posible. Una decisión que se debe tomar es si conviene usar números o letras para la notación. Si se opta por los números, una base de un dígito podrá albergar hasta 10 facetas (del 0 al 9). Para extenderla sería necesario recurrir a una base de dos dígitos (del 00 al 99) o de tres dígitos (del 000 al 999), lo cual puede producir notaciones muy extensas. Si se opta por letras, la base de una sola letra podría albergar hasta 26 facetas y sería más fácil mantener una notación breve.
- **Hospitalidad:** el sistema de notación debe permitir que en el futuro los nuevos conceptos sean acomodados en facetas, subfacetas o focos, sin alterar las notaciones existentes.
- **Expresividad:** la notación debe evidenciar el orden sistemático del tesoro.

Según los símbolos utilizados, la notación puede ser de dos tipos (Gil Urdiciain, 2004):

- **Notación pura:** utiliza un solo tipo de símbolo, por ejemplo solo números o solo letras.
- **Notación mixta:** utiliza una combinación de símbolos, por ejemplo letras y números, o letras mayúsculas y letras minúsculas.

Respecto al orden, también hay dos tipos de notación (Vickery, 1960a):

- Notación jerárquica: indica claramente el orden del esquema de lo general a lo particular. Tiene la ventaja de ser expresiva, pero también la desventaja de perder brevedad.
- Notación ordinal: ordena los conceptos en el esquema, pero no hace evidente las jerarquías. No es expresiva, pero ayuda a mantener la brevedad.

En la Tabla 1 se compara un ejemplo de notación ordinal con otro de notación jerárquica. Nótese que ambos tipos de notación expresan el orden del esquema, pero mientras que en la notación jerárquica es fácil de reconocer (la notación más corta representa un concepto más general y la notación más larga representa un concepto más específico), en la notación ordinal no es posible distinguir qué conceptos están subordinados a otros.

Tabla 1. Comparación de una notación ordinal y una jerárquica

Ordinal	Concepto	Jerárquica
a	Sudamérica	1
b	Argentina	11
c	Chaco	111
d	Resistencia	111.1
e	Barrio España	111.11
f	Corrientes	112
g	Misiones	113
h	Brasil	12

Autoevaluación

¿Qué se entiende por dominio del tesoro?

¿En qué consisten los ejes Áreas de Modulación y Grados de Especialización?

¿Qué es el análisis de dominio?

¿Qué se entiende por temas marginales?

¿Cuáles son las etapas del análisis de facetas?

¿Qué son las categorías?

¿Cuál es la diferencia entre aislado y foco?

¿Qué es una faceta?

Defina atributo, característica de división y subfaceta.

¿Cuáles son los principios del análisis de facetas?

¿Qué otros tipos de orden sistemático acepta la norma ISO 25964?

¿Qué es la notación y para qué sirve en un tesoro?

Describa las cualidades que debe poseer la notación.

Evaluación, publicación y actualización del tesoro

Los procesos finales del diseño y desarrollo de un tesoro son la evaluación, la publicación y la actualización. Los explicaremos a continuación.

Evaluación del tesoro

La evaluación es el proceso de obtener evidencias para juzgar el grado en que se han alcanzado los objetivos y las metas establecidas. En consecuencia, la evaluación del tesoro consiste en obtener evidencias para juzgar en qué grado cumple con las funciones de indización y recuperación.

Como afirman Owens y Cochrane (2004), la evaluación de tesoros puede realizarse por distintos métodos. Por ejemplo, puede ser evaluado por un experto, comparado con otros tesoros, contrastado con las normas, probado en experiencias reales de indización y recuperación, etc. Con distinta terminología, Gil Leiva (2008) y Lancaster (2002) reconocen varios tipos de evaluación de tesoros, que se presentan en la Tabla 10.1.

Tabla 1. Métodos de evaluación de tesoros

Tipo de evaluación	Metodología	
Evaluación intrínseca o superficial	Cualitativa	Se evalúa la introducción, el uso de calificadores y notas, reciprocidad de las relaciones, formas gramaticales según las normas, etc.
	Cuantitativa o estadística	Evaluación por medio de indicadores
Evaluación extrínseca	Comportamiento del tesoro en la indización y la recuperación	Prueba de indización, tasa de precisión y tasa de acierto

Evaluación intrínseca cualitativa

La evaluación intrínseca cualitativa requiere que un experto o un grupo de expertos revise las características del tesauro y las contraste contra la norma que se ha seguido. Esta evaluación debe determinar si el tesauro consta de una portada, una introducción con todos los elementos que se mencionan en el Apartado 10.2.1, si incluye calificadores y notas, si la reciprocidad de las relaciones es consistente, si las formas gramaticales, la tipografía y las etiquetas utilizadas son las recomendadas, etc.

Evaluación intrínseca cuantitativa

La evaluación intrínseca cuantitativa o estadística requiere la aplicación de indicadores.

La experiencia más conocida para la evaluación de tesauros usando indicadores es la realizada por el Bureau Marcel Van Dijk (BMD) de Bruselas, que en 1976 llevó a cabo un estudio sobre 10 tesauros monolingües y 10 tesauros multilingües (Commission des Communautés Européennes, 1976). El BMD propuso varios indicadores que también han sido aplicados por otros autores (Álvaro Bermejo *et al.*, 1989a,b), aunque Lancaster (2002) llegó a la conclusión de que si bien algunos son ingeniosos, otros resultan triviales e incluso arbitrarios. Por otro lado, hay que tener en cuenta que el BMD basó su evaluación en la primera edición de la norma ISO 2788:1974, que fue sustituida por una segunda edición hace casi 30 años y recientemente por la norma ISO 25964.

Nosotros hemos propuesto un conjunto de indicadores, basados en las normas publicadas en este siglo veintiuno. Aunque fueron probados en una muestra de tres tesauros, la intención ha sido que sean incorporados al software de gestión, de modo que puedan producir informes estadísticos automáticamente. Los indicadores se describen a continuación (Martínez Tamayo *et al.*, 2009, 2011a):

a. Porcentaje de términos preferentes que designan más de un concepto; por ejemplo, el término preferente nombres y códigos identificadores de observación lógica del *Tesaurus Spines*. Se define como $A/B \times 100$, donde A es el número de términos preferentes que designan más de un concepto y B es el número total de términos preferentes. El valor de referencia es 0, ya que las normas establecen que un término preferente debe designar un solo concepto.

b. Media de palabras por término preferente. Se define como A/B , donde

A es el número de palabras de los términos preferentes y B es el número total de términos preferentes. Se expresa como $\pm DE$ (media \pm un desvío estándar). El valor de referencia es $2,0 \pm 1,0$, es decir, un rango promedio entre 1 y 3 palabras por término preferente. Este rango coincide con el estudio realizado por Gil Leiva & Rodríguez Muñoz (1997), quienes para una muestra de 1284 términos preferentes asignados en bases de datos científicas españolas, identificaron un 86% con una, dos o tres palabras.

c. Porcentaje de términos preferentes con formas gramaticales no especificadas por las normas; por ejemplo, el término preferente información/biblioteca, cooperación del Tesoro de la Unesco. Se define como $A/B*100$, donde A es el número de términos preferentes que presentan formas gramaticales distintas a las especificadas por la norma y B es el número total de términos preferentes. El valor de referencia debería ser 0, pero como también aquí las normas admiten excepciones, se debe asumir que mientras menor sea este porcentaje, mayor será el cumplimiento con las recomendaciones de la norma.

d. Porcentaje de términos preferentes con formas de singular y plural no recomendados. Se define como $A/B*100$, donde A es el número de términos preferentes que presentan formas de singular o plural distintas a las especificadas por las normas y B es el número total de términos preferentes. El valor de referencia es 0, ya que en este caso no hay excepciones.

e. Porcentaje de términos preferentes que carecen de relación jerárquica (términos huérfanos). Se define como $A/B*100$, donde A es el número de términos preferentes que carecen de relación jerárquica y B es el número total de términos preferentes. El valor de referencia para este indicador es 0, ya que todo término preferente debe tener al menos una relación jerárquica.

f. Porcentaje de términos preferentes con dos o más términos genéricos. Se define como $A/B*100$, donde A es la suma de términos preferentes que poseen dos o más TG y B es el número total de términos preferentes. Si se aplica el concepto de estructura monojerárquica, entonces el valor de referencia de este indicador es 0, ya que ningún término preferente puede tener más de un TG. Si se aplica el concepto de estructura polijerárquica (véase el Apartado 8.1.4) no hay un valor de referencia, pero siendo que la norma considera que esta situación se produce solo en algunos casos y debe tener sólidos fundamentos lógicos, se asume que mientras menor sea el porcentaje, mayor será el cumplimiento con la norma.

g. Porcentaje de términos preferentes con un solo término específico. Se define como $A/B*100$, donde A es el número de términos preferentes que poseen un solo TE y B es el número total de términos preferentes. El valor de referencia para este indicador es 0, ya que en un tesoro no puede haber un término preferente con un solo TE, según el principio de diferenciación postulado por Ranganathan (véase el Apartado 9.4).

Evaluación extrínseca

Finalmente, para la evaluación extrínseca que mide el comportamiento del tesoro en la indización y la recuperación se recomienda lo siguiente:

- Prueba de indización: consiste en distribuir una edición preliminar del tesoro entre un cierto número de bibliotecas especializadas, centros de documentación u otras instituciones de la BDCI para una experiencia real de indización, durante un período de varios meses. Las instituciones participantes deben comunicar todos los errores y fallas que encuentren, así como las sugerencias que estimen convenientes, tanto del propio tesoro como de la interfaz.
- Prueba de recuperación: las mediciones más conocidas son las tasas de precisión y de acierto propuestas por Claverdon y Mills (1985). En el resultado de estas mediciones intervienen distintas variables como la exhaustividad y la especificidad de la indización, la interfaz usuario-sistema, la experiencia de búsqueda de los usuarios finales, etc., pero también pueden aplicarse específicamente a la evaluación del tesoro, cuando la recuperación se realiza exclusivamente por los términos preferentes y las formas autorizadas del nombre (Foskett, 1996; Lancaster, 1995).

La tasa de precisión ha sido incorporada como indicador en la norma IRAM-ISO 11620 (IRAM, 2007). Se define como A/B , donde A es el número de documentos recuperados que son relevantes y B es el número total de documentos recuperados en la búsqueda. Es una medida de la capacidad del sistema de recuperación para encontrar un documento por un punto de acceso determinado, en este caso un término preferente o una forma autorizada del nombre.

La tasa de acierto se define como A/B , donde A es el número de documentos relevantes recuperados y B es el número total de documentos relevan-

tes en la base de datos. Es una medida de la capacidad del sistema de recuperación para reunir conjuntos de documentos que respondan a un determinado punto de acceso, por ejemplo, un término preferente o una forma autorizada del nombre.

Entre la tasa de precisión y la de acierto existe una relación inversamente proporcional, es decir, a mayor precisión, menor acierto y a mayor acierto, menor precisión. En general, se privilegia la precisión.

Los tipos de evaluación que se han mencionado no son excluyentes, sino complementarios: se pueden aplicar todos o algunos al mismo tesoro y de hecho conviene que así se haga.

Cuándo realizar la evaluación

Hay dos momentos en los que debe haber una evaluación del tesoro por parte del equipo de trabajo responsable de su diseño y desarrollo:

- Una primera evaluación antes de publicar la primera edición, como se indicó más arriba.
- Periódicamente para su actualización, teniendo en cuenta además las sugerencias, comentarios y quejas que los indizadores y los usuarios finales hayan hecho llegar.

Los métodos de evaluación sugeridos aquí también se pueden aplicar a un tesoro cuando se necesita considerar su adquisición, pero esta evaluación obviamente es realizada por terceros. Resulta muy útil que quienes efectúan este tipo de evaluaciones hagan conocer los resultados a los miembros del equipo de trabajo responsable del tesoro.

Publicación del tesoro

La publicación del tesoro puede ser impresa o digital. Hasta hace algunos años, la mayoría de los tesoros se publicaban impresos en papel, pero en la actualidad se cuenta con numerosos tesoros digitales, muchos de ellos disponibles en la Web. La tendencia, por supuesto, es propiciar la incorporación de tesoros en línea y que esta sea la forma de publicación habitual. La norma ISO 25964 refleja esta tendencia, en contraposición con las normas anteriores que ponían el énfasis en la publicación impresa (ISO, 2011).

Cualquiera sea el tipo de publicación, los componentes del tesoro son:

- Portada o página de inicio: identifica al tesoro con la mención del título, los nombres de los creadores y colaboradores, el número de edición, la editorial, el lugar y la fecha de publicación, el ISBN, DOI, etc.
- Tabla de contenido: debe enumerar todas las secciones del tesoro y remitir a estas mediante hipervínculos. Si el tesoro está disponible en la Web, la tabla de contenido puede presentarse en una tabla lateral o en solapas.
- Introducción: debe brindar, tanto al indizador como al usuario final, la información que se detalla en el Apartado 10.2.1.
- Sección sistemática: como su nombre lo indica, despliega el orden sistemático del tesoro. Debe estar hipervinculada con la sección alfabética.
- Sección alfabética: muestra los términos preferentes y no preferentes, así como las formas autorizadas y variantes del nombre, en una secuencia alfabética. Debe estar hipervinculada con la sección sistemática.

Introducción del tesoro

La introducción del tesoro tiene como propósito informar, tanto al indizador como al usuario final, los aspectos que se mencionan a continuación, en lo posible con abundantes ejemplos ilustrativos:

- El dominio del tesoro, su alcance y los temas marginales. Conviene mencionar a qué tipo de institución, usuarios finales y documentos está destinado.
- Los idiomas en que se presenta.
- Indicación de la norma que se ha seguido para desarrollarlo.
- El significado de todas las convenciones, abreviaturas, etiquetas y signos de puntuación empleados en él.
- El número de términos y nombres propios, discriminados por términos preferentes y no preferentes, formas autorizadas y variantes.
- La fecha de incorporación del último término o nombre propio.
- Las reglas o criterios empleados en la selección de los conceptos, la selección y redacción de los términos y nombres propios.
- La política de actualización, incluyendo la frecuencia, las fechas y los procedimientos pertinentes, así como el nombre y domicilio del responsable a donde deben enviarse los comentarios y sugerencias.
- Referencias a las fuentes utilizadas en la compilación del tesoro, así

como los agradecimientos a las personas y organizaciones ajenas al equipo multidisciplinario que han colaborado con el tesauro. En este sentido, es necesario contar con el permiso escrito de las personas u organizaciones a mencionar en tales agradecimientos.

Registros individuales de términos y nombres propios

Para cada término preferente o forma autorizada del nombre se debe generar un registro individual, que debe mostrar los siguientes datos en el mismo orden que se presenta:

término preferente o forma autorizada del nombre

CC Notación o código de clasificación

NA Nota de alcance

UP *término no preferente*

TT término tope

TG término genérico

TE término específico

TR término relacionado

DEF Nota de definición

NH Nota histórica

CM Categoría a la que se asigna el concepto

Tomando nuevamente el ejemplo del capítulo 9, se tendría:

profesionales

CC 220

NA Úsese para obras que tratan sobre los profesionales universitarios en general. Para obras sobre los profesionales que ejercen una profesión específica, úsese el término que designe esa profesión, por ejemplo, bibliotecarios, informáticos.

UP *profesionistas*

TT personal

TG personal

TE bibliotecarios

TE informáticos

TE licenciados

TE magisters

DEF Miembros del personal de la biblioteca, cuya función requiere del ejercicio de una profesión universitaria.

NH Incorporado el 30-11-2013

CM Agentes

Por otra parte, el tesoro debe desplegar también las relaciones de equivalencia desde un término no preferente o una forma variante del nombre al respectivo término preferente o forma autorizada del nombre. Por ejemplo:

Profesionistas
USE profesionales

Drabenstott, Karen
USE Markey, Karen

Tipografía

Existe cierta libertad en cuanto a la tipografía que debe emplearse, pero se recomienda diferenciar claramente los términos preferentes de los no preferentes y las formas autorizadas del nombre de sus formas variantes. No obstante, la norma ISO 25964 sugiere las siguientes pautas (véanse los ejemplos del Apartado 10.2.2):

- El término preferente y la forma autorizada del nombre se deben escribir con letra regular (ni cursiva, ni negrita).
- El término no preferente y la forma variante del nombre se escriben con letra cursiva.
- Las etiquetas de las relaciones (USE/UP, TG/TE, TR/TR) de las notas, de la notación y de la categoría de materia se indican en mayúscula compacta y negrita.
- La etiqueta de nodo, como ya se mencionó, se indica en letra cursiva y entre corchetes. Cuando se refiere a la característica de división, se antepone la preposición *por*; no así cuando se trata de la categoría. Por ejemplo:
Categoría: *[agentes]*

Característica: *[por función]*

- Los términos preferentes y no preferentes se escriben en minúsculas, incluyendo su letra inicial. Se exceptúan aquellos casos que incluyen

obligatoriamente una letra mayúscula. Por ejemplo:

biólogos

D-alanina transaminasa

novelas policiales

pH

En las formas autorizadas y variantes del nombre se usará mayúscula solo para las letras iniciales, como indica la gramática española; el resto de las letras en minúscula. Por ejemplo:

Misiones (Argentina)

Torre Nilsson, Leopoldo, 1924-1978

Universidad de Buenos Aires

- No conviene usar las mayúsculas compactas o en bloc para resaltar el término preferente o la forma autorizada del nombre porque dificultan la lectura; por ejemplo:

plaguicidas y no PLAGUICIDAS

- Las siglas se exceptúan de la instrucción anterior, ya que deben escribirse en mayúscula compacta y sin punto, de modo que puedan recuperarse como una palabra:

CONICET

SIDA

UNESCO

- Cuando los términos van acompañados de números, letras griegas o signos de puntuación, estos deben respetarse. Las letras griegas se escriben como una palabra para facilitar la indización y la recuperación. Por ejemplo:

cadena beta de alfa-cristalina

molécula 1 de adhesión intercelular

N-acetil muramoil-L-alanina amidasa

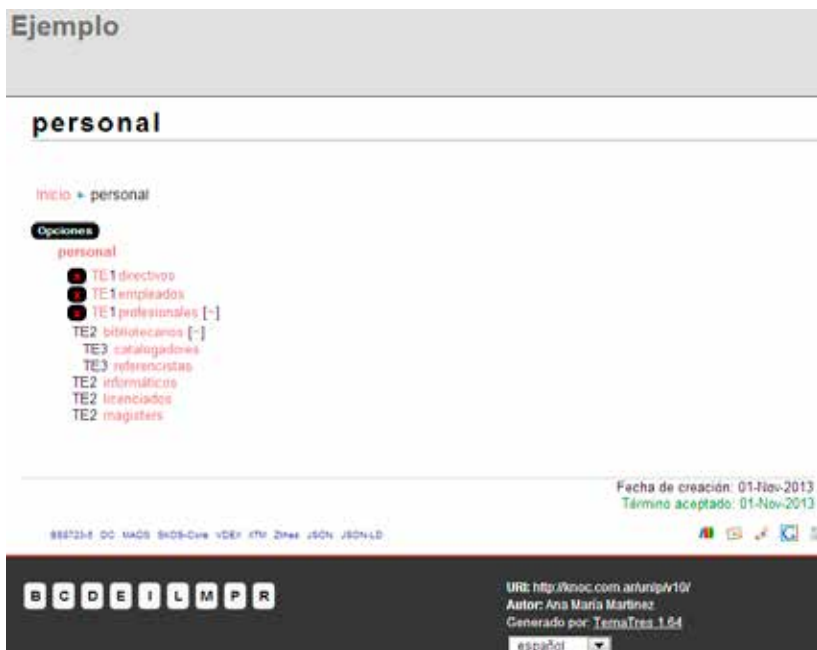
receptor erbB-3
xilano endo-1,3-beta-xilosidasa

Secciones sistemática y alfabética

En la Figura 1 se muestra un ejemplo de la sección sistemática correspondiente al ejemplo que hemos venido desarrollando.

La sección sistemática no requiere indicar las relaciones de equivalencia o las relaciones asociativas, sino solamente el orden sistemático completo que incluye las relaciones jerárquicas entre términos preferentes o entre estos y las formas autorizadas de los nombres (Aitchison *et al.*, 2000; Gil Urdaín, 2004; ISO, 2011; Lancaster, 2002; Naumis Peña, 2008).

Figura1. Ejemplo de la sección sistemática del tesoro. La jerarquía se muestra con TE1, TE2, TE3. Además, cada nivel de la jerarquía se puede expandir [+] o contraer [-]. La interfaz corresponde al software TemaTres.



Una forma novedosa de presentar la sección sistemática del tesoro es a través de un mapa de tópicos como el que se muestra en la Figura 2.

[illegible]

Por su parte, la sección alfabética, como su nombre lo indica, muestra los términos preferentes y no preferentes, más las formas autorizadas y variantes de los nombres, siguiendo la secuencia alfabética. Bajo cada término preferente o cada forma autorizada del nombre se debe desplegar toda la información incluida en el registro individual, como se mencionó en el Apartado 10.2.2. A su vez, bajo cada término no preferente y cada forma variante del nombre debe aparecer el reenvío al término preferente o a la forma autorizada del nombre que corresponda.

¹ Consulte la versión interactiva en <http://highwire.stanford.edu/hjt/HWTree.html?catid=0>.

Figura 3. Ejemplo de registro individual de un término preferente. La flecha indica la cadena **Inicio** „ **personal** „ **profesionales** „ **bibliotecarios** „ **catalogadores**. Esta cadena pertenece a la faceta personal. Nótese también la nota de alcance, el término genérico y el término relacionado. La interfaz corresponde al software TemaTres.

Ejemplo

catalogadores

Inicio ▶ personal ▶ profesionales ▶ bibliotecarios ▶ catalogadores

Nota de alcance:

Bibliotecarios que se especializan en el proceso de catalogación.

TG bibliotecarios

catalogadores

TR catalogación

Fecha de creación: 01-Nov-2013
Término aceptado: 01-Nov-2013

658723-5 DC MAD5 SKOS-Core VDEX XTM Zthes JSON JSON-LD

BCDEILMPR

URI: <http://knoc.com.ar/unlp/v10/>
Autor: Ana María Martínez
Generado por: TemaTres 1.64

español

Un tesauo debe vincular la parte alfabética con la sistemática, de modo que tanto el indizador como el usuario final puedan comprobar qué lugar ocupa un término o un nombre propio en el sistema de conceptos representado en él. En un tesauo impreso era obligatorio que cada término preferente o cada forma autorizada del nombre tuvieran una notación, porque esta servía de nexo entre ambas secciones. En cambio, en un tesauo electrónico el nexa entre las dos secciones está dado por los hipervínculos o enlaces correspondientes. Por otro lado, algunos tesauos impresos incluían un índice permutado para facilitar la búsqueda de un término o nombre propio particular. En un tesauo digital esto tampoco es necesario, porque se sustituye por el casillero de búsqueda, que puede localizar cualquier palabra sin importar si se trata de la

primera o alguna posterior.

En el Apéndice B de este capítulo se muestran varios ejemplos de tesauros electrónicos.

Actualización del tesoro

Como ya se mencionó, el tesoro necesita una revisión permanente de los términos y nombres propios y del ajuste correspondiente de las relaciones y la ubicación en el orden sistemático, con el fin de agregar nuevos términos, eliminar los obsoletos y modificar los que exijan determinados cambios, según sea necesario. El tesoro digital y en línea tiene la gran ventaja de que las actualizaciones pueden estar disponibles en forma inmediata para indizadores y usuarios finales, o al menos acortar los tiempos de actualización. Es de gran utilidad para el equipo de trabajo ofrecer la posibilidad tanto a los indizadores como a los usuarios finales que utilicen el tesoro, de enviar sus críticas, comentarios o sugerencias, para lo cual se puede incluir un formulario especial. De esta manera, se contará con un valioso aporte para la futura actualización. El equipo de trabajo debe mantenerse activo, más allá de la publicación de la primera edición, para actualizarla periódicamente.

Autoevaluación

¿En qué consiste la evaluación del tesaurus?

Explique los tipos de evaluación que se han descripto para el tesaurus.

¿En qué momento debe evaluarse el tesaurus?

¿Qué secciones debe contener un tesaurus?

Indique los aspectos que no deben faltar en la introducción de un tesaurus.

¿Qué tipografía sugiere la norma de tesauros para los términos, nombres propios y etiquetas de relación?

¿Qué información debe presentarse bajo un término preferente o una forma autorizada del nombre?

¿Qué información se incluye en la sección alfabética del tesaurus?

¿Qué información se incluye en la sección sistemática del tesaurus?

Apéndice B: Ejemplos de tesauros electrónicos

Figura B1. Interfaz de búsqueda del Tesauro AGROVOC.²



The screenshot shows the AGROVOC Thesaurus search interface. At the top, there is a navigation bar with the AIMS logo and the text 'Agricultural Information Management Standards'. Below this, there is a search bar with the text 'Search Terms in the AGROVOC Thesaurus'. To the left of the search bar, there is a sidebar with various links including 'Home', 'News & Events', 'Open Access', 'Vocabularies', 'AGROVOC', 'AgroTagger', 'VocBench', 'VIST: Vocabularies', 'Metadata', 'Tools', 'Linked Data', and 'Community'. The main search area includes a 'Language' dropdown menu with options for English, French, Spanish, Arabic, and Chinese. Below this, there is a 'Search in:' dropdown menu with options for Sub-vocabularies, All terms, Acronyms, Chemicals, Common name for animals, and Common name for bacteria. There are also checkboxes for 'Show:' (Current Terms, Deleted Terms, All Terms) and radio buttons for search criteria (starting with, end with, containing text, exact match, exact word). The search bar contains the text 'maíz'.

El tesauro es publicado por la Organización de las Naciones Unidas para la Alimentación y la Agricultura. Nótese el carácter multilingüe del mismo, que obliga a marcar el idioma (Spanish). También puede elegirse el tipo de término que se desea buscar (todos los términos, acrónimos, compuestos químicos, nombres comunes de animales y bacterias, etc.). Se puede buscar por los términos en uso o por aquellos que han sido eliminados. Permite elegir si la expresión de búsqueda utilizada es la palabra inicial, final, contenida en el texto o exacta. Finalmente, se encuentra el casillero de búsqueda en el que se ha indicado *maíz* como término a buscar

² <http://aims.fao.org/es/agrovoc>

Figura B2. Lista alfabética de términos del Tesauro AGROVOC que incluyen la palabra *maíz*.

Results from Term Description:

- Use the  button to see the hierarchy of the corresponding term
- Use the  button to see more information on the corresponding term.

Code	Term	Status	Scope
12332	Maíz (ES)	Descriptor	 
7152	Maíz amiláceo (ES)	Descriptor	 
22788	Maíz arrugado (ES)	Non-Descriptor	 
10110	Maíz azucarado (ES)	Non-Descriptor	 
23878	Maíz blando (ES)	Non-Descriptor	 
12337	Maíz cereo (ES)	Non-Descriptor	 
8337	Maíz ceroso (ES)	Descriptor	 
10104	Maíz córneo (ES)	Non-Descriptor	 
10111	Maíz de cera (ES)	Non-Descriptor	 
11739	Maíz de la India (ES)	Non-Descriptor	 
6108	Maíz de palomitas (ES)	Descriptor	 
22789	Maíz de rosetas (ES)	Non-Descriptor	 
22790	Maíz de rositas (ES)	Non-Descriptor	 
2107	Maíz dentado (ES)	Descriptor	 
7552	Maíz dulce (ES)	Descriptor	 

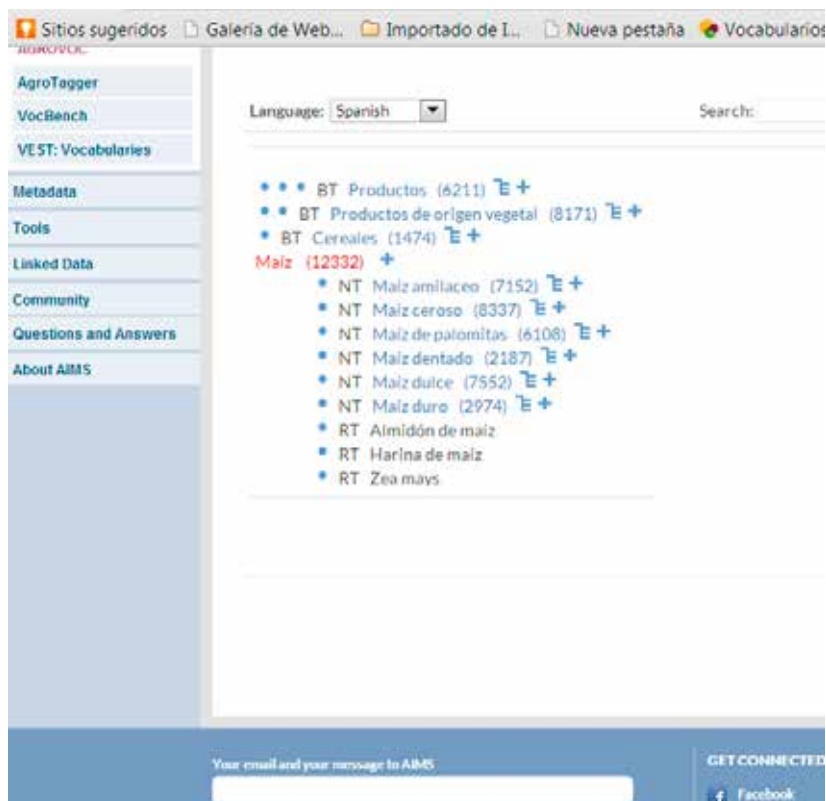
Nótese lo siguiente: a) en la primera columna aparece el código o notación de cada término; b) en la segunda columna, los términos ordenados alfabéticamente; c) en la tercera columna, el estatus del término (descriptor o no descriptor); y d) en la cuarta columna aparecen dos opciones: el botón + permite ver el registro individual del término preferente, mientras que el botón E remite al orden sistemático. Los términos *descriptor* y *no descriptor*, de larga historia, se encuentran muy difundidos. Sin embargo, la norma ISO 25964 ha considerado más adecuados *término preferente* y *término no preferente*, respectivamente.

Figura B3. Registro individual de un término preferente en el Tesouro AGROVOC.

Vocabularies	Labels	Status	Scope	Created	Last modified
Metadata	Maiz (ES)	Descriptor (20)	n/a	1987-01-22	1987-01-22 00:00:00
Tools					
Linked Data	Word Tree				
Community	UF	25456 - this term does not exist in ES			
Questions and Answers	BT	1474 - Cereales (ES)			
About AIMS	NT	14385 - this term does not exist in ES			
	NT	8337 - Maiz ceroso (ES)			
	NT	7552 - Maiz dulce (ES)			
	NT	7152 - Maiz amilaceo (ES)			
	NT	6108 - Maiz de palomitas (ES)			
	NT	2974 - Maiz duro (ES)			
	NT	2187 - Maiz dentado (ES)			
	RT	25472 - Harina de maiz (ES)			
	RT	15500 - Almidón de maiz (ES)			
	RT	8504 - Zea mays (ES)			
	Descriptions				
	No descriptions found for the selected term.				

En la parte superior aparece el término preferente en español Maíz (ES); a continuación, el estatus indicando que se trata de un descriptor o término preferente, no tiene nota de alcance (SN) y las notas históricas indican la fecha de alta y la última modificación, que en este caso coinciden. Luego se despliegan todas las relaciones: a) UF (used for) para la relación de equivalencia, con la aclaración de que no hay equivalencias en español; b) BT (broader term o término genérico) que en este caso es Cereales (ES); c) luego aparecen varios NT (narrower terms o términos específicos) y finalmente d) los RT (related terms o términos relacionados). No hay descripciones adicionales para este término preferente, pero podría haberlas para otros términos.

Figura B4. Despliegue del orden sistemático del Tesauro AGROVOC.



Hacia arriba del término preferente Maíz se incluye el término genérico inmediato, en este caso Cereales (ES), y los términos superiores a este. Cada punto indica un nivel más alto de especificidad en el orden sistemático. Por debajo del término preferente Maíz se incluyen los términos específicos (NT) y relacionados (RT). Los términos genéricos y específicos ofrecen la posibilidad de marcar el botón **E** para continuar navegando por el orden sistemático, o bien la opción **+** para ir directamente al registro individual del término.

Figura B5. Interfaz de búsqueda del Tesauro SPINES.³

Este tesauro ha sido elaborado por la Unesco y su versión en línea y en español es gestionada por el Instituto de Estudios Documentales sobre Ciencia y Tecnología del Consejo Superior de Investigaciones Científicas de España. Como se puede apreciar, cuenta con un casillero de búsqueda para localizar el término que interesa al usuario. También se pueden marcar, en las pestañas que aparecen en la parte superior de la interfaz, tanto la sección alfabética como la sección sistemática o jerárquica.

³ http://thes.cindoc.csic.es/buscar_form_esp.php?thes=SPIN

Figura B6. Sección sistemática del Tesauro Spines.

The screenshot displays the 'Tesauro de Spines' web application. At the top, there is a navigation bar with links for 'Contacto', 'Tarifas', a search box with a 'Buscar' button, and 'Inicio' with a flag icon. Below this, a secondary navigation bar lists 'Tesauros', 'Inicio', 'Jerárquico', 'Alfabético', and 'Búsquedas'. The left sidebar contains several menu categories: 'El CINDOC' (Presentación, Memoria), 'Investigación' (Proyecto, Resultados), 'Productos' (Bases de Datos CSIC, Monografías, Tesauros Impresos), and 'Acceso Gratuito' (Sumarios de Revistas, Directorios, Revistas del IEDCYT, Portales Científicos, Tesauros en Línea, Enlaces). The main content area is titled 'Tesauro de Spines' and includes a note: 'Los términos que van acompañados de asterisco (*) no deben usarse como descriptores en la indicación'. Below this, a hierarchical tree structure is shown, starting with 'Tesauros.' and branching into various fields like 'Política, Gobierno, Administración Pública', 'Derecho, Legislación', 'Economía, Desarrollo económico, social', and 'Desarrollo'. The 'Desarrollo' branch is expanded, showing sub-levels like 'AUTOSUFICIENCIA', 'CAMBIO SOCIOCULTURAL', 'CAMBIO TECNOLÓGICO', 'DESARROLLO ECONÓMICO', and further down, 'DESARROLLO AGRÍCOLA', 'DESARROLLO COMERCIAL', 'DESARROLLO INDUSTRIAL', 'CAMBIO INTERINDUSTRIAL', and 'INDUSTRIALIZACIÓN'.

Se puede navegar a través del orden sistemático del tesauro abriendo o cerrando los botones + o -. Así, la faceta de primer nivel Economía, Desarrollo económico y social puede abrirse en un segundo nivel de especificidad que incluye, entre otros, el término preferente Desarrollo. Este a su vez puede abrirse en un tercer nivel de especificidad en el que se encuentra Desarrollo industrial. Finalmente, este se abre en un cuarto y último nivel que incluye dos términos preferentes: Cambio interindustrial e Industrialización. No todas las jerarquías cuentan con cuatro niveles, ya que puede haber algunas con menos o más niveles de especificidad. Un clic en cualquiera de los términos remite al registro individual del término preferente.

Figura B7. Sección alfabética del Tesauro SPINES.

INSTITUTO DE ESTUDIOS DOCUMENTALES
SOBRE CIENCIA Y TECNOLOGÍA
(antes CSICCI)
Consejo Superior
de Investigaciones Científicas

Contato Tarfas [Inicio](#)  English

Tesauros Inicio Jerárquico Alfabético Búsquedas

EI IEDCYT
Presentación
Memoria

Investigación
Proyectos
Resultados

Productos
Bases de Datos CSIC
Micro-SIS
Tesauros Impresos

Acceso Gratuito
Sumarios de Revistas
Directorios
Revistas del IEDCYT
Portales Científicos
Tesauros en Línea
Enlaces

Tesauro SPINES

1 2 A B C D E F G H I J K L M N Ñ O P Q R S T U V W X Y Z

Listado alfabético de términos [F10] (no-descriptores en cursiva)
[Inicio] [Anterior] 5 6 7 8 9 **10** 11 12 13 14 15 [Siguiente] [Fin]

DE SARROLLO DE TEORÍAS
DE SARROLLO DE TESAURUS
DESARROLLO DEL ROL
DE SARROLLO ECONÓMICO
DESARROLLO ECONÓMICO Y SOCIAL
DE SARROLLO EMERSIONADO
DE SARROLLO ENDOGENO
DE SARROLLO EXPERIMENTAL
DESARROLLO EXPLORATORIO
DE SARROLLO FÍSICO
DE SARROLLO INDIVIDUAL
DE SARROLLO INDUSTRIAL
DESARROLLO INTEGRAL
DESARROLLO INTELECTUAL
DE SARROLLO INTERNACIONAL

Nótese en la parte superior el índice A/Z. La lista muestra una parte de la página 10 de la letra D, en la que se encuentra el término preferente Desarrollo industrial. Los términos en negrita son los preferentes, mientras que los no preferentes se indican en letra cursiva. También aquí un clic en cualquiera de los términos remite al registro individual del término preferente.

Figura B8. Registro individual de un término preferente en el Tesauro SPINES.

The screenshot displays the Tesauro SPINES website interface. At the top, the header includes the CSIC logo, the text 'INSTITUTO DE ESTUDIOS DOCUMENTALES SOBRE CIENCIA Y TECNOLOGIA (antes CINACAL) Consejo Superior de Investigaciones Científicas', and a search bar with a 'Buscar' button. Navigation links for 'Inicio' and 'Inglés' are present. Below the header, a green navigation bar contains links for 'Tesauros', 'Inicio', 'Jerárquico', 'Alfabético', and 'Búsquedas'. The main content area is titled 'Tesauro SPINES' and features a sidebar on the left with categories like 'El IEDCYT', 'Investigación', 'Productos', and 'Acceso Gratuito'. The central part of the page lists terms under the heading 'DE DESARROLLO INDUSTRIAL'. The term 'DE DESARROLLO INDUSTRIAL' is highlighted in blue. Below it, several related terms are listed, including 'TG DESARROLLO ECONOMICO', 'TE CAMBIO INTERINDUSTRIAL', 'TE INDUSTRIALIZACIÓN', 'TR CAMBIO EN LA TÉCNICA', 'TR CONSULTORES INDUSTRIALES', 'TR DESARROLLO COMERCIAL', 'TR DIVERSIFICACIÓN DE LA PRODUCCIÓN', 'TR ECONOMÍA INDUSTRIAL', 'TR I+D INDUSTRIAL', 'TR INDUSTRIAS BASADAS EN LA CIENCIA', 'TR PLANIFICACIÓN INDUSTRIAL', 'TR POLÍTICA INDUSTRIAL', 'TR POTENCIAL INDUSTRIAL', 'TR PRODUCCIÓN INDUSTRIAL', 'TR PRODUCTOS INDUSTRIALES', 'TR PROYECTOS INDUSTRIALES', 'TR SECTOR INDUSTRIAL', 'EN INDUSTRIAL DEVELOPMENT', and 'FR DÉVELOPPEMENT INDUSTRIEL'.

El término preferente Desarrollo industrial se resalta en color azul. Se incluye un término genérico (TG) y dos términos específicos (TE), así como numerosos términos relacionados (TR). Aunque no han sido necesarios en este caso, el tesauro también muestra las notas de alcance (NA) y los términos equivalentes (UP). Por último, reenvía a los equivalentes en inglés y francés, ya que se trata de un tesauro multilingüe.

Figura B9. Interfaz del Tesauro de OIT.⁴



Este tesauro es gestionado por la Organización Internacional del Trabajo. Nótese el índice A/Z que constituye la sección alfabética del tesauro. Debajo de este índice aparecen tres opciones: términos tope, huérfanos (que no deberían existir) y las categorías por las que se ha analizado el dominio. El tesauro es multilingüe, de modo que en el casillero de búsqueda se ha indicado el idioma para buscar el término Bibliotecario. Más abajo se puede ver la leyenda “Resultados de búsqueda para Bibliotecario”

⁴ <http://www.ilo.org/thesaurus/defaultes.asp>

Figura B10. Registro individual de un término preferente en el Tesoro de OIT.

The screenshot displays the OIT Thesaurus website. At the top, there is a navigation bar with links: 'Agencia de la OIT', 'Temas', 'Regiones', 'Eventos y reuniones', 'Publicaciones', 'Investigación', 'Normas del trabajo', 'Estadísticas y bases de datos'. Below this is the header 'INFORM Oficina de Biblioteca y Servicios de Información' and 'Departamento de Comunicación y de la Información Pública'. The main content area is titled 'Tesauro OIT' and features a search bar with the text 'Términos que principian con...' and a dropdown menu showing 'bibliotecario'. Below the search bar, there is a grid of letters (A-Z) and a section for 'Términos top', 'Huellas', and 'Categorías'. The main entry for 'bibliotecario' is displayed, showing its classification (SC: 12.11.2 Profesiones, técnicas y trabajos liberales), its generic term (BT: trabajador profesional), related terms (RT: biblioteca, bibliotecología, documentalista), and equivalents in French (FRE: bibliothécaire) and English (ENG: librarian). A sidebar on the left contains links to 'La biblioteca de la OIT', 'Recursos de información en línea', and 'Herramientas para especialistas de la información'. A sidebar on the right contains 'Links' and a '¿Alguna pregunta?' button. The footer indicates the last update on 12/12/2011.

Nótese la categoría a la que pertenece el término (SC), el término genérico (BT), los términos relacionados (RT) y las equivalencias en francés (FRE) e inglés (ENG). Por último se muestra la nota histórica (última actualización).

Figura B11. Interfaz del *Tesaurus europeo de educación*.⁵



Es gestionado por la Comisión de las Comunidades Europeas. Directamente, la página de inicio proporciona la lista sistemática, que muestra las distintas categorías bajo las cuales se han ordenado los términos preferentes. En forma horizontal se dispone de la lista alfabética A/Z. En la parte superior, aparece el casillero de búsqueda y se brinda acceso a la búsqueda avanzada y a la información del tesaurus.

⁵ <http://www.vocabularyserver.com/tee/es/index.php>

Figura B12. Registro individual de un término preferente en el *Tesaurus europeo de educación*.



Nótese en la parte superior la cadena correspondiente al término. Luego, en la pestaña Término, aparecen el término preferente resaltado en negrita, los términos genéricos (TG), los términos específicos (TE), los términos relacionados (TR) y las equivalencias en los otros idiomas del tesaurus. Las notas y los metadatos se visualizan en pestañas separadas.

El tesoro y la interoperabilidad

En la actualidad y gracias a la Web, existe la posibilidad de que un sistema de recuperación pueda incorporar una función de búsqueda distribuida, es decir, una búsqueda simultánea en múltiples fuentes, cualquiera sea su localización. Sin embargo, este tipo de búsqueda por lo general se encuentra limitada a la lengua natural, de modo que el usuario enfrenta los problemas de sinonimia y homonimia que se explicaron en el capítulo 2. A esto debe agregarse la dificultad que presenta la búsqueda distribuida cuando el contenido de los documentos se encuentra en distintas lenguas, como ocurre en la Web.

Por otra parte, muchos sistemas de recuperación también requieren intercambiar los registros bibliográficos de bases de datos desarrolladas previamente con total independencia. Tales bases de datos suelen estar indizadas con distintos SOC, de modo que necesitan compatibilizarse, tanto en lo conceptual como en lo terminológico, mediante alguna forma de interoperabilidad (Martínez Tamayo *et al.*, 2011b; Neelamegan & Lalitha, 2013).

La interoperabilidad se define como “la habilidad de dos o más sistemas para intercambiar información y usar la información que ha sido intercambiada” (ISO, 2013: 7). Así, la interoperabilidad entre SOC se propone armonizar las equivalencias y las relaciones jerárquicas y asociativas entre los conceptos de dos o más SOC, con el fin de intercambiar esos conceptos. Un tesoro puede interoperar con otro tesoro, con una lista de epígrafes, un sistema de clasificación o cualquier otro SOC (Aitchison *et al.*, 2000; Chang & Zeng, 2002a,b; Lancaster, 2002; Martínez Tamayo *et al.*, 2011b; Neelamegan & Lalitha, 2013).

Esta compatibilidad entre SOC no es una novedad. Ya en 1939 la conocida lista de encabezamientos de materia de Sears asignó a cada epígrafe una notación de SCDD, haciendo ambos SOC compatibles (Miller & Bristow, 2007). En las décadas de 1970 y 1980 se llevaron a cabo diversos proyectos

de este tipo (Aitchison *et al.*, 2000; Lancaster, 2002). Para 1995 el tema había adquirido tal importancia que la ISKO organizó en Varsovia el *Seminario sobre Compatibilidad e Integración de Sistemas de Ordenación*, que culminó con una serie de recomendaciones tanto para la investigación, la educación y el intercambio de información, como para el desarrollo de sistemas y esfuerzos cooperativos (ISKO, 1996).

En este siglo veintiuno, la disponibilidad de tesauros y otros SOC en línea, así como los recursos que ofrece la tecnología informática, han facilitado numerosos y variados proyectos de interoperabilidad entre SOC. En consecuencia, las nuevas normas de diseño y desarrollo de tesauros han dedicado buena parte de sus recomendaciones a la interoperabilidad (BSI Group, 2005-2008; ISO, 2013; NISO, 2010).

Modelos estructurales de interoperabilidad

La norma ISO 25964-2 describe distintos modelos de interoperabilidad entre SOC que, si bien se presentan por separado para facilitar la explicación, en la práctica suelen superponerse o combinarse (ISO, 2013).

Unidad estructural: en este modelo, todos los SOC participantes comparten exactamente los mismos conceptos y las mismas relaciones jerárquicas y asociativas. En cambio, los términos pueden variar, por ejemplo, cuando se trata de dos o más SOC en distintos idiomas.

Vinculación directa: este modelo vincula dos o más SOC que no comparten la misma estructura. Además de diferenciarse en el alcance, el idioma y la estructura, la vinculación puede incluir distintos tipos de SOC, por ejemplo un tesoro con un sistema de clasificación o con una lista de epígrafes. Debe establecerse un mapeo directo entre los conceptos de un SOC y aquellos de cada uno de los otros SOC. El modelo puede extenderse a cualquier cantidad de SOC y todos los SOC pueden usarse para la búsqueda en una colección indizada con cualquiera de los otros SOC (Figura 1).

Estructura central o *hub*: en lugar de establecer un mapeo directo entre cada par de SOC, se elige un SOC como *hub* o SOC central, hacia el cual se mapea cada uno de los otros SOC, de modo que cada concepto en el *hub* debe ser mapeado al correspondiente concepto de los otros SOC y viceversa, pero no entre todos los SOC. La interoperabilidad se establece entre cada SOC y el *hub* (Figura 2).

Figura 1. Modelo de interoperabilidad de vinculación directa. Cada uno de los SOC vinculados es compatible con todos y cada uno de los demás SOC.

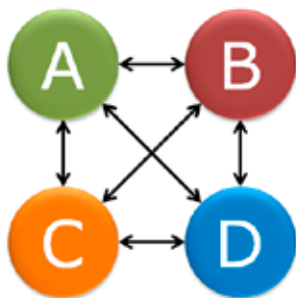
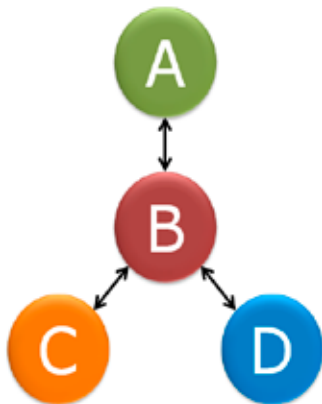
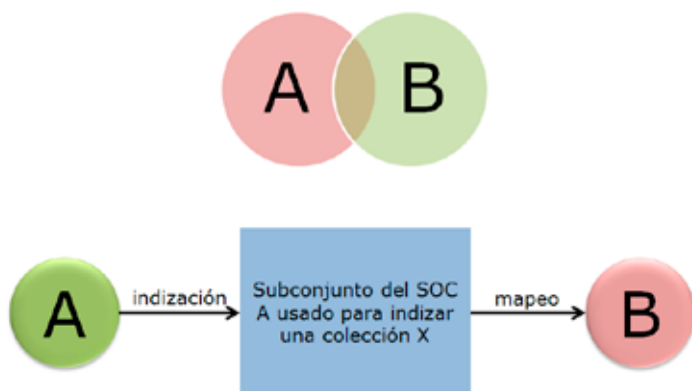


Figura 2. Modelo de interoperabilidad de estructura de *hub*. El SOC B es el *hub*. Cada uno de los otros SOC se vincula con el *hub*, pero no con los demás SOC. Es decir, para vincular el SOC A con el SOC C, es necesario pasar a través del SOC B que es el *hub*.



Mapeo selectivo: en algunos casos no es necesario mapear la totalidad de los SOC, sino solamente una parte de ellos. Hay dos opciones: mapear el área que se superpone o mapear solamente los términos preferentes que se han usado en la indización (Figura3).

Figura 3. Modelo de interoperabilidad de mapeo selectivo. Arriba: solo se mapean los conceptos superpuestos o que forman la intersección del SOC A y del SOC B. Abajo: se mapean los términos del SOC A usados realmente en la indización y los términos del SOC B.



El primer modelo —el de unidad estructural— se recomienda para un tesauruso multilingüe nuevo, donde cada concepto es designado por un término preferente en cada una de las lenguas y donde las relaciones jerárquicas y asociativas son las mismas. Por el contrario, los demás modelos se recomiendan para mapear SOC existentes y de distinto tipo, porque al ser diseñados independientemente, los conceptos y relaciones pueden no coincidir de manera exacta.

El método para lograr la interoperabilidad es el mapeo, es decir, el proceso de establecer relaciones entre los conceptos de un SOC y los conceptos de otro SOC. El mapeo es más complicado cuando los SOC involucrados son de diferente tipo, particularmente entre un SOC poscoordinado y un SOC precoordinado, ya que en este último es necesario descomponer primero las cadenas precoordinadas.

La interoperabilidad ocurre entonces entre dos o más SOC, que pueden ser de dos tipos según su función:

- SOC fuente: es el SOC que sirve como punto de partida cuando se busca un término o concepto correspondiente en otro SOC.
- SOC blanco: es el SOC en el que se encuentra un término o concepto correspondiente a un término o concepto del SOC fuente.

Mapeo equivalente

El punto crítico de la interoperabilidad es la equivalencia, pues no es posible establecer jerarquías y asociaciones si no se han establecido primero las equivalencias entre términos. Se debe diferenciar entre a) la equivalencia intra-SOC, que se explicó en el capítulo 7 y b) la equivalencia inter-SOC que se describe aquí.

Esta equivalencia inter-SOC se establece entre las designaciones de los distintos SOC, es decir, entre los términos preferentes y las formas autorizadas del nombre. Por ejemplo, en el tesoro A el término preferente es batata y en el tesoro B es boniato, o bien en el Tesoro A la forma autorizada del nombre es Cortázar, Julio y en el Tesoro B es Cortázar, Julio, 1914-1984. También pueden estar en distinta lengua, por ejemplo guarderías y crèches, Suiza y Suisse.

En otros casos el concepto de un SOC se designa con un término, mientras que en el otro SOC se designa con dos o más. Por ejemplo en el Tesoro A el término preferente es políticas de salud y en el Tesoro B se requieren dos términos preferentes: políticas + salud.

Un término puede representar un concepto en un tesoro y otro concepto diferente en un segundo tesoro. Por ejemplo, un tesoro de fotografía puede incluir el término preferente diafragma para referirse al dispositivo que regula la cantidad de luz que ha de pasar en las cámaras fotográficas, mientras que un tesoro de acústica puede tener el término preferente diafragma para referirse al dispositivo que transforma las vibraciones del sonido en impulsos eléctricos o viceversa en algunos aparatos de sonido como el micrófono.

Hay entonces distintos tipos de equivalencia inter-SOC, que se indican con la etiqueta EQ:

- Equivalencia simple: el SOC blanco contiene un concepto idéntico al concepto del SOC fuente, permitiendo un inmediato mapeo uno-a-uno. Un ejemplo: el término preferente enseñanza secundaria del Tesoro de la Unesco y la notación 373.5 de CDU. Otros ejemplos:

SOC fuente		SOC blanco
jardinería	EQ	635
computadoras	EQ	computadores
computadoras	EQ	ordenadores
maní	EQ	amendoim
Países Bajos	EQ	Deutschland

En la equivalencia simple se distinguen dos grados de equivalencia exacto e inexacto que se indican con los símbolos = y ~, respectivamente.

Equivalencia exacta: en un mapeo equivalente, la equivalencia es exacta cuando el término preferente del SOC fuente y el término preferente del SOC blanco designan el mismo concepto y pueden usarse indistintamente. Por ejemplo: castañas de cajú =EQ anacardos.

Equivalencia inexacta: en un mapeo equivalente, los términos preferentes considerados equivalentes no siempre son exactamente lo mismo. La equivalencia entonces es inexacta. Por ejemplo: bibliotecología ~EQ bibliotecnología.

- Equivalencia compuesta: un concepto complejo que ha sido designado en el SOC fuente por un único término preferente puede estar expresado en el SOC blanco por una combinación de dos o más términos. Se produce así un mapeo uno-a-muchos. Por ejemplo: el epígrafe de la lista AR-MARC administración de fábricas es equivalente a la precoordination de dos términos preferentes del Tesauro de la Unesco: administración + fábrica. Cuando en un mapeo selectivo se convierten términos de indización, todos los términos preferentes del SOC blanco deben aplicarse al registro indizado con el término del SOC fuente. En la búsqueda deben combinarse con el operador booleano AND. El símbolo usado es la adición (+). Por ejemplo:

SOC fuente		SOC blanco
políticas de salud	EQ	políticas + salud
economía doméstica	EQ	economía + hogar
ciencia y tecnología	EQ	ciencia + tecnología
gestión de museos	EQ	gestión + museos
autos de competencia	EQ	autos + competencias

- Equivalencia compuesta por intersección: debe su nombre a que es posible representarla gráficamente como una intersección de conjuntos. Ocurre cuando el concepto del SOC blanco tiene un alcance más general que el concepto en el SOC fuente. Se establece una relación del tipo envío ascendente. Por ejemplo, el término preferente isomerasa de peptidilprolil del DeCS y el término preferente isomerasas del Tesauro Spines,

que no cuenta con términos más específicos. Este tipo de equivalencia no es reversible, pues el término general también puede abarcar otros conceptos específicos. Por ejemplo:

SOC fuente		SOC blanco
bibliotecas públicas	EQ	bibliotecas
papa blanca	EQ	papas
perros de caza	EQ	perros
automóviles deportivos	EQ	automóviles
Jujuy	EQ	Argentina

- Equivalencia compuesta acumulada: es aquella en la que un concepto complejo del SOC fuente equivale a la suma o conjunto de dos o más conceptos del SOC blanco. Por ejemplo: el término preferente corrientes de agua del Tesoro Spines equivale a la suma de los epígrafes aguas | canales | ríos | cataratas (hidrografía) de la lista ARMARC. Cuando en un mapeo selectivo se convierten términos de indización, todos los términos preferentes del SOC blanco deben aplicarse al registro indizado con el término preferente del SOC fuente. En la búsqueda deben combinarse con el operador booleano OR. El símbolo de la barra vertical (|) debe indicarse entre los términos preferentes del SOC blanco.

SOC fuente		SOC blanco
transportes	EQ	aviones barcos trenes automotores
escuelas	EQ	escuelas públicas escuelas privadas
jóvenes	EQ	mujeres jóvenes varones jóvenes
frutos secos	EQ	nueces castañas avellanas maníes
flores	EQ	flores de jardín flores silvestres

Mapeo jerárquico

El mapeo jerárquico se establece cuando un concepto es claramente más general que el otro. Al igual que la relación jerárquica en un tesoro, la relación entre los conceptos de distintos SOC puede ser del tipo género-especie o enumerativa. La relación todo-parte también puede mapearse jerárquicamente, en las circunstancias limitadas que describe la norma ISO 25964-1.

El mapeo jerárquico de un concepto específico a un concepto genérico se expresa con el formato: amebas BM protozoos. El mapeo jerárquico recíproco es protozoos NM amebas. BM significa *broader mapping* o mapeo genérico, mientras que NM significa *narrower mapping* o mapeo específico. Dado que no existe todavía una norma de interoperabilidad entre tesauros en español, en esta obra se ha decidido mantener las etiquetas en inglés.

Mapeo asociativo

El mapeo asociativo se establece cuando los conceptos de dos o más SOC no califican para el mapeo equivalente o jerárquico, pero están asociados semánticamente. La etiqueta para este mapeo es RM (*related mapping*). Por ejemplo, alcohol RM alcoholismo y alcoholismo RM alcohol. Al igual que en el caso anterior, dado que no existe todavía una norma de interoperabilidad entre tesauros en español, en esta obra se ha decidido mantener las etiquetas en inglés.

La norma ISO 25964-2 aclara que la línea divisoria entre el mapeo asociativo y la equivalencia inexacta es difícil de definir y subjetiva, por lo cual recomienda hacerlo pragmáticamente de acuerdo con el contexto en el que se usará el mapeo.

Tesauros multilingües

Un tesoro multilingüe es aquel en el que cada concepto debe estar representado por igual en todas las lenguas incluidas, es decir, todos los idiomas son tratados en igualdad de condiciones. Ejemplos de este tipo son los tesauros del sistema de la ONU, entre ellos el Tesoro de la Unesco, el Tesoro de la OIT, el de UNBIS, AGROVOC, etc.

La metodología para tratar un tesoro multilingüe es una forma de interoperabilidad. Las recomendaciones específicas provienen de dos fuentes: la norma ISO 25964-1/2 (ISO, 2011; 2013) y las directrices de IFLA para tesauros multilingües (IFLA, 2009a).

Cuando todos los conceptos del tesoro multilingüe se expresan con un término en cada idioma y comparten las relaciones jerárquicas y asociativas, se dice que es un tesoro simétrico. Lo ideal por supuesto es un tesoro multilingüe simétrico, pero no siempre es posible, ya que puede haber variaciones de un idioma a otro y de hecho se han identificado varios problemas que

suelen presentarse. En tal caso, se trata de un tesauruso asimétrico.

Antes de continuar, es necesario definir dos términos:

- Lengua fuente: es la lengua que sirve como punto de partida en la traducción de un término o en la búsqueda de sus equivalencias.
- Lengua blanco: es la lengua que proporciona la traducción o el equivalente de un término que existe en una lengua fuente.

Problemas frecuentes

A continuación se presentan los problemas más comunes en los tesaurusos multilingües. En los ejemplos, la lengua se indica en negrita y de acuerdo al código de dos letras para idiomas de la norma ISO 639-1: en = inglés, es = español, fr = francés, it = italiano, pt = portugués (ISO, 2002).

El término en la lengua fuente posee un calificador que no es necesario en la lengua blanco:

Lengua fuente	en cranes (birds) en cranes (lifting equipment)
Lengua blanco	es grúas es grullas

- El término en la lengua fuente tiene sinónimos, pero no en la lengua blanco. Se respeta la situación sin imponer sinónimos en el tesauruso de la lengua blanco.

Lengua fuente	es computadoras UP <i>computadores</i> UP <i>ordenadores</i> <i>computadores</i> USE computadoras <i>ordenadores</i> USE computadoras
---------------	--

Versión en inglés	tomato es tomate it pomodoro UF <i>Solanum lycopersicum</i>
Versión en español	tomate en tomato it pomodoro UP <i>jitomate</i> UP <i>Solanum lycopersicum</i>
Versión en italiano	pomodoro en tomato es tomate UP <i>Solanum lycopersicum</i>

Nótese que la versión en español del ejemplo anterior tiene un sinónimo más (jitomate) que las otras versiones. Esto es así porque la cantidad de sinónimos y cuasi sinónimos de un término varía de una lengua a otra. En un tesoro multilingüe no se requiere la equivalencia entre los términos no preferentes de las distintas lenguas.

Algunas experiencias de interoperabilidad con tesauros

Sin pretender ser exhaustivos, a continuación se presentan algunas experiencias de interoperabilidad entre uno o varios tesauros, o entre un tesoro y otro tipo de SOC.

Entre dos o más tesauros

- CAT-AGROVOC: este proyecto compatibiliza el Tesoro Chino de Agricultura (CAT) y el tesoro multilingüe AGROVOC de la FAO (Liang *et al.*, 2005).
- ASEG: es un proyecto del Centro de Ecosistemas Acuáticos del Instituto de Investigación y Tecnología Agropecuaria de Cataluña. La interoperabilidad se ha establecido entre cinco tesauros que abarcan las ciencias acuáticas, pesca, agricultura, alimentación, medio ambiente y

economía (Reverté & Sebastián, 2009).

- Unified Astronomy Thesaurus (UAT): es un proyecto coordinado por la American Astronomy Society para crear un tesoro colaborativo con el fin de compatibilizar varios tesauros independientes de astronomía y astrofísica. En el proyecto participan asociaciones profesionales, sociedades científicas, editores, bibliotecarios y curadores con experiencia en el dominio y en tesauros adicionales (Accomazzi *et al.*, 2014).
- Thesaurus of GTS: es un proyecto realizado por profesionales de China y los Países Bajos, quienes han creado un tesoro multilingüe en siete idiomas sobre las escalas de los tiempos geológicos (Ma *et al.*, 2011).

Entre tesauros y listas de epígrafes

- LCSH-ERIC: es un trabajo experimental realizado en OCLC para probar la metodología de trabajo Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). La metodología se aplicó al tesoro de ERIC y a la lista de epígrafes LCSH (Vizine Goetz *et al.*, 2004). Esta metodología también ha sido descrita en Argentina (Ferreira, 2009).
- SWD-IZ: el objetivo de esta experiencia alemana ha sido relacionar los conceptos indizados en dos bases de datos, una que indiza con el Thesaurus Sozialwissenschaften de ciencias sociales y otra que utiliza la lista de epígrafes Schlagwortnormdatei (SWD) (Zhang, 2006).

Entre tesauros y sistemas de clasificación

- Proyecto HILT. Es un proyecto llevado a cabo por el Centre for Digital Library Research de la University of Strathclyde en Gran Bretaña. El objetivo es lograr un mapeo de diversos tesauros en inglés, usando como *hub* SCDD (Mcgregor *et al.*, 2007; McCulloch & Macgregor, 2008; Nicholson, 2002; Nicholson & McCulloch, 2006).
- Proyecto KoMoHe: es un proyecto del Bundesministerium für Bildung und Forschung de Alemania, iniciado en 2004 y concluido en 2007. En total se vincularon 25 SOC (16 tesauros, 4 listas de términos, 2 listas de epígrafes y 3 sistemas de clasificación, incluyendo SCDD), de 11 disciplinas, en alemán, inglés y ruso (Mayr & Petras, 2008, 2009).
- Multilingual Thesaurus of Education: es un proyecto piloto realizado en la India. Vincula un tesoro en inglés, derivado del esquema de edu-

cación de la Colon Classification (CC), y diversas lenguas locales (tamil, kannada, tulu, malayalam, punjabi, hindi, sánscrito), a las cuales se piensa agregar el urdu y el árabe. El proyecto incluye además las notaciones de CC y SCDD (Neelamegan & Lalitha, 2013).

- ERIC-TEE-CDU: es un estudio experimental realizado en la Universidad de Barcelona, que estableció la equivalencia entre los términos preferentes del tesauro ERIC en inglés y el Tesauro Europeo de Educación (versión en catalán), con notaciones de CDU (Granados Colillas & Nicolau, 2007, 2008). Se trata de un mapeo selectivo, basado en la indización.
- ASC-CDU: es un proyecto realizado en la biblioteca del Afrika-Studiecentrum (ASC) de Leiden, Holanda, entre 2001 y 2006. Cada término preferente del tesauro de ASC tiene asignada una notación de CDU, con lo cual se abre la posibilidad de establecer más adelante la conexión con otros tesauros, usando CDU como *hub* (Doorn & Polman, 2010).
- BiblioPhil: es un proyecto de interoperabilidad realizado en la Biblioteca Centrală Universitară de Bucarest, Rumania, que ha derivado un tesauro trilingüe (rumano, inglés y francés) a partir de las clases de CDU (Frâncu & Sabo, 2010). Los términos preferentes en las tres lenguas quedan así vinculados a las notaciones.
- STITCH. Entre 2006 y 2007, las bibliotecas nacionales de Holanda y Francia (Koninklijke Bibliotheek o KB y Bibliothèque Nationale de France o BNF) llevaron a cabo este proyecto con el fin de hacer interoperables el sistema de clasificación Iconclass, con el que se clasifica la base de datos de manuscritos medievales de la KB, y el tesauro Mandragore con el que se indiza otra base de datos sobre el mismo tema en la BNF (Angeli & Isaac, 2008a,b).
- VISION: es un sistema experimental desarrollado en China, que consiste en una red integrada formada por la fusión de un tesauro y un sistema de clasificación en un nuevo SOC denominado Clasificación y Tesauro Chinos. Este SOC ha sido aplicado a los registros bibliográficos de la biblioteca de la Universidad de Beijing (Wang, 2003).

Tesauros y ontologías

Algunos tesauros han sido usados como base para crear ontologías y por este motivo se consideran lenguajes de transición. Las relaciones en una on-

tología son más formales que en un tesoro, y en general se considera que el tesoro requiere una reingeniería tanto estructural como conceptual para ser reusado o tratado como una ontología, pero puede ser la base para su desarrollo (Kless *et al.*, 2012).

Dificultades de la interoperabilidad entre SOC

Lancaster (2002) ha señalado los principales problemas de la interoperabilidad entre SOC:

1. Nivel de especificidad
2. Nivel de precoordinación
3. Solapamiento de términos
4. Falta de normalización
5. Grado de actualización de los SOC

Para solucionar los primeros tres problemas, Tennis (2004) ha propuesto focalizar la investigación en el concepto como unidad de conocimiento, la materia como cuerpo de ideas organizado o sistematizado y el orden sistemático facetado. Con respecto a la normalización, hoy es un problema resuelto con la nueva norma de diseño y desarrollo de tesauros ISO 25964 (ISO, 2011, 2013). En cuanto a la actualización, es obvio que si uno de los SOC que interopera no se actualiza más o menos al mismo tiempo que los otros, será muy difícil mantenerlos interoperables.

Además, existen algunas dificultades metodológicas, pues si bien los métodos informáticos resultan efectivos cuando se trata de mapear términos en el área de las ciencias duras, en otros campos —como las ciencias sociales— se necesita recurrir al trabajo intelectual de los expertos para establecer el correspondiente mapeo (Lauser *et al.*, 2008).

Autoevaluación

En el contexto de los SOC, defina el término *interoperabilidad*.

Explique los distintos modelos de interoperabilidad.

¿A qué se llama SOC fuente y SOC blanco?

Señale cómo se establece el mapeo equivalente.

¿Qué etiqueta indica una equivalencia inter-SOC?

Describa los distintos tipos de equivalencias inter-SOC.

Señale cómo se establece el mapeo jerárquico.

¿Qué etiquetas indican una relación jerárquica inter-SOC?

Señale cómo se establece el mapeo asociativo.

Explique por qué un tesoro multilingüe se debe tratar como una forma de interoperabilidad.

¿Qué problemas se pueden presentar en el diseño de un tesoro multilingüe?

Software de gestión, modelo de datos y formatos de intercambio

En capítulos anteriores hemos hecho hincapié en la necesidad de presentar los tesauros en línea. Este capítulo está dedicado a las recomendaciones básicas de la norma ISO 25964-1/2 para el *software* de gestión de la base de datos del tesoro, el modelo de datos y los formatos de intercambio, así como a algunas referencias al desarrollo automático de tesauros (ISO, 2011, 2013).

Software de gestión

Un tesoro debe gestionarse como una base de datos. Hay programas adecuados para este fin, como [TemaTres](#), que ha sido desarrollado en Argentina y aplicado a más de cien SOC en distintos idiomas. Otros *software* de este tipo han sido evaluados y comentados en la Web.¹

También se puede diseñar la base de datos del tesoro utilizando los siguientes formatos de autoridades:

- [MARC 21 format for authority data](#), desarrollado por Library of Congress.
- [UNIMARC manual authority format](#), desarrollado por IFLA.

Ambos formatos incluyen los campos necesarios para el término preferente, las relaciones de equivalencia, jerárquicas y asociativas, las equivalencias en otras lenguas, así como la notación de un sistema de clasificación si se desea hacerlo compatible con el tesoro (Library of Congress, 2013; Willer, 2009).

Los requisitos funcionales de FRSAD también aportan lo necesario para desarrollar una lista de autoridades de materia, que puede aplicarse perfectamente al diseño de un tesoro. Como se mencionó en el capítulo 3, FRSAD es un modelo conceptual para datos de autoridades de materia, desarrollado por IFLA y adecuado para implementar bases de datos relacionales, que ha

¹ <http://www.willpowerinfo.co.uk/thessoft.htm>

establecido dos superclases *thema* y *nomen* cuyos atributos y relaciones son los siguientes (IFLA, 2010):

- *Thema*: los atributos son el tipo de *thema* y la nota de alcance; las relaciones son las jerárquicas y las asociativas.
- *Nomen*: los atributos son tipo de *nomen*, SOC, fuente de referencia, representación, idioma, escritura del *nomen*, escritura de conversión, forma del *nomen*, tiempo de validez, audiencia y estatus. La relación es la de equivalencia.

La norma ISO 25964 describe las recomendaciones básicas para el programa informático de gestión del tesoro, que pueden resumirse como sigue (ISO, 2011):

- Capacidades: el programa debe contar con capacidad de almacenamiento, gestión, importación, exportación y edición de datos, así como de búsqueda, visualización y navegación. Es conveniente que pueda gestionar más de un tesoro, en uno o más idiomas. Debe permitir la incorporación, modificación y eliminación de términos y nombres propios candidatos, términos preferentes y no preferentes, formas autorizadas y variantes del nombre, notas, notaciones, relaciones semánticas y niveles jerárquicos.
- Limitaciones: el programa no debe tener limitaciones respecto al número de caracteres de los términos y nombres propios, notas, etiquetas, etc. Tampoco debe limitar la cantidad de términos, de relaciones por término o de los niveles jerárquicos, y debe facilitar la inclusión de facetas, subfacetas y etiquetas de nodo. Debe establecer automáticamente las reciprocidades correspondientes a las relaciones semánticas del tesoro.
- Validación: el programa debe validar todas las relaciones de modo que no haya en el tesoro ni términos huérfanos, ni relaciones sin la debida reciprocidad. También debe verificar que cada concepto esté representado por un solo término preferente o una sola forma autorizada del nombre, así como la ortografía.
- Vinculación: es indispensable que el programa permita vincular la base de datos del tesoro con la base de datos bibliográfica que será indizada y consultada con dicho tesoro, de modo tal que tanto el indizador como el usuario final puedan operar con ambas bases de datos simultáneamente.

También se recomienda que sea compatible con los modelos de datos

apropiados para incluir un tesoro en la Web, como SKOS o MADS.

Modelo de datos

Un modelo de datos es un lenguaje que describe representar y usar los datos en una base de datos. En el caso del tesoro, proporciona la descripción genérica de su estructura y su semántica.

La norma ISO 25964-1 recomienda un modelo diseñado de acuerdo con las convenciones del lenguaje unificado de modelado (UML o *unified modelling language*). Contiene cinco clases principales que son: Thesaurus, ThesaurusArray, ThesaurusConcept, ThesaurusTerm y Note, así como otras clases subsidiarias de estas. También indica el carácter obligatorio u opcional de los datos y sus posibles ocurrencias (repetibles, no repetibles). El modelo UML propuesto por ISO y su correspondiente esquema XML se pueden encontrar en el sitio que NISO ha dedicado a esto: <http://www.niso.org/schemas/iso25964/schema-intro/>.

Formatos de intercambio

Además del modelo propuesto por la norma ISO, deben tenerse en cuenta los formatos de intercambio como el [Metadata authority description schema](#) (MADS) de Library of Congress y el [Simple Knowledge Organization System](#) (SKOS), que es un formato elaborado por el W3C para la Web Semántica, es decir, para ordenar el caos de la Web .

SKOS ha sido desarrollado específicamente para expresar cualquier SOC incluidos los tesauros de forma tal que sus datos puedan intercambiarse entre distintas aplicaciones informáticas y publicarse en la Web (W3C, 2009). **Además, SKOS se ha mapeado con el modelo FRSAD de IFLA (Zeng & Zimmer, 2009a,b).** En SKOS los datos se expresan como tripletas RDF (sujeto-predicado-objeto) y pueden codificarse usando cualquier sintaxis RDF, por ejemplo RDF/XML. El *Resource Description Framework* (RDF) es un marco de especificaciones para representar información en la Web. El *eXtensible Markup Language* (XML) es un lenguaje de marcado para almacenar datos e intercambiar información estructurada.

En SKOS, un SOC cualquiera es concebido como un esquema de conceptos. Tanto el esquema de conceptos como cada uno de los conceptos que lo integran se reconocen por un identificador uniforme de recursos (URI o

Uniform Resource Identifier), que consiste en una cadena de caracteres que identifica inequívocamente un recurso Web, de modo que cualquiera pueda referirse al esquema o a cada uno de los conceptos sin ambigüedad. Este URI facilita que los conceptos SKOS puedan reusarse en otros esquemas y que se facilite el mapeo entre dos o más esquemas, por lo que SKOS resulta ideal para la interoperabilidad entre SOC.

A cada concepto SKOS se le puede asignar una o más *etiquetas léxicas*, en cualquier idioma o escritura. También se pueden asignar *notaciones* y documentar *notas* de alcance, de definición, históricas o de edición. Un concepto SKOS puede vincularse a otro concepto SKOS mediante relaciones semánticas de equivalencia, jerarquía y asociación. Además, los conceptos SKOS pueden agruparse en *colecciones* como las etiquetas de nodo o las categorías.

El vocabulario de SKOS se resume en la Tabla 1. En la Figura 1 se presenta un ejemplo tomado del *Tesaurus de la Unesco*, marcado con SKOS.

Tabla 1 Resumen del vocabulario del modelo SKOS

Vocabulario SKOS	Descripción	
Conceptos y esquemas de conceptos	skos:Concept skos:ConceptScheme skos:inScheme skos:hasTopConcept	Concepto Esquema de conceptos Concepto en un esquema Concepto tope
Etiquetas léxicas	skos:prefLabel skos:altLabel skos:hiddenLabel	Término preferente Término no preferente Término oculto
Relaciones semánticas	skos:broader skos:narrower skos:related skos:broaderTransitive skos:narrowerTransitive	Concepto genérico Concepto específico Concepto relacionado Relación jerárquica transitiva Relación jerárquica transitiva
Notas documentadas	skos:note skos:scopeNote skos:historyNote skos:changeNote skos:definition skos:editorialNote skos:example	Nota Nota de alcance Nota histórica Nota por modificación Nota de definición Nota de los editores Nota de ejemplo

Colecciones	skos:Collection skos:OrderedCollection skos:member skos:memberList	Colección de conceptos Colección ordenada de conceptos Concepto en una colección Concepto en una colección ordenada
-------------	---	--

Figura 1. Modelo de concepto SKOS tomado del Tesauro de la Unesco.² Se ha resaltado lo siguiente: en rojo el concepto (skos:Concept) con su correspondiente URI. En azul el término preferente (skos:prefLabel) en los cuatro idiomas del tesauro: inglés, francés, español y ruso. En lila las relaciones jerárquicas (skos:narrower y skos:broader). En verde las relaciones asociativas (skos:related). En violeta las relaciones de equivalencia (skos:altLabel) en los cuatro idiomas.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
ns#" xmlns:skos="http://www.w3.org/2004/02/skos/core#">
  <skos:Concept rdf:about="http://skos.um.es/unescothes/C00525">
    <skos:prefLabel xml:lang="en">Cataloguing</skos:prefLabel>
    <skos:prefLabel xml:lang="fr">Catalogage</skos:prefLabel>
    <skos:prefLabel xml:lang="es">Catalogación</skos:prefLabel>
    <skos:prefLabel xml:lang="ru">Каталогизирование</skos:prefLabel>
    <skos:inScheme rdf:resource="http://skos.um.es/unescothes/C00000"/>
    <skos:narrower rdf:resource="http://skos.um.es/unescothes/C00367"/>
    <skos:broader rdf:resource="http://skos.um.es/unescothes/C01985"/>
    <skos:related rdf:resource="http://skos.um.es/unescothes/C00374"/>
    <skos:related rdf:resource="http://skos.um.es/unescothes/C00524"/>
    <skos:related rdf:resource="http://skos.um.es/unescothes/C00526"/>
    <skos:related rdf:resource="http://skos.um.es/unescothes/C01085"/>
    <skos:related rdf:resource="http://skos.um.es/unescothes/C01088"/>
    <skos:related rdf:resource="http://skos.um.es/unescothes/C02003"/>
    <skos:altLabel xml:lang="en">Bibliographic description</skos:altLabel>
    <skos:altLabel xml:lang="en">Cataloging</skos:altLabel>
    <skos:altLabel xml:lang="en">Centralized cataloguing</skos:altLabel>
    <skos:altLabel xml:lang="en">Cooperative cataloguing</skos:altLabel>
    <skos:altLabel xml:lang="en">Descriptive cataloguing</skos:altLabel>
    <skos:altLabel xml:lang="en">Metadata</skos:altLabel>
    <skos:altLabel xml:lang="es">Catalogación centralizada</skos:altLabel>
    <skos:altLabel xml:lang="es">Catalogación descriptiva</skos:altLabel>
    <skos:altLabel xml:lang="es">Descripción bibliográfica</skos:altLabel>
    <skos:altLabel xml:lang="fr">Catalogage centralisé</skos:altLabel>
    <skos:altLabel xml:lang="fr">Catalogage coopératif</skos:altLabel>
    <skos:altLabel xml:lang="fr">Catalogage descriptif</skos:altLabel>
    <skos:altLabel xml:lang="fr">Description bibliographique</skos:altLabel>
    <skos:altLabel xml:lang="fr">Metadata</skos:altLabel>
    <skos:altLabel xml:lang="ru">Библиографическое описание</skos:altLabel>
    <skos:altLabel xml:lang="ru">Централизованное
    каталогизирование</skos:altLabel>
    <skos:altLabel xml:lang="ru">Объединенное
    каталогизирование</skos:altLabel>
    <skos:altLabel xml:lang="ru">Описательное
    каталогизирование</skos:altLabel>
    <skos:altLabel xml:lang="ru">Метаданные</skos:altLabel>
  </skos:Concept>
</rdf:RDF>
```

²<http://skos.um.es/unescothes/C00525/rdfxml>

Desarrollo automático de tesauros

En cuanto al aporte informático para el desarrollo de tesauros, se distinguen tres metodologías:

- Tesauro manual: desde luego no se trata de un tesauro hecho u operado con la mano, como hace suponer el adjetivo. Es un tesauro cuyo diseño y desarrollo se basa principalmente en el trabajo intelectual de seres humanos, expertos en el dominio del tesauro y en los métodos para su elaboración. La tecnología informática es usada como una herramienta de apoyo que se aplica a los procesos involucrados, como la colecta de términos u otros. Desde el sector informático, especialmente de los desarrolladores Web, se considera que este tipo de tesauros resulta caro, consume tiempo, se desactualiza y es ineficiente para la búsqueda en la red (Zeng, 2012), aunque por supuesto esto es muy discutible.

- Tesauro automático: es un tesauro desarrollado por computadoras programadas especialmente. Se trata de una estructura de datos construida mediante la extracción de palabras claves y la determinación automática de las relaciones entre esas palabras claves. El investigador chino Wen Zeng considera que el desarrollo de un tesauro automático tiene cuatro etapas (Zeng, 2012):

- a. Extracción de términos. Incluye tanto la extracción propiamente dicha de los términos de los documentos, especialmente del área científica y técnica, como el análisis lingüístico de esos términos, en particular las estructuras sintácticas.

- b. Determinación automática de las relaciones semánticas. Para las relaciones asociativas se indica un análisis estadístico de coocurrencias o una asociación basada en la estructura de los documentos. Para las relaciones jerárquicas se toma en cuenta primero la forma de los términos y luego se calcula por similitud, de acuerdo al número de palabras similares y a la localización del término en el documento.

- c. Traducción automática de los términos. Tanto Zeng como otros

autores consideran indispensable que los tesauros operativos en la Web sean multilingües para poder realizar una búsqueda exitosa en ella, dada la enorme cantidad de recursos disponibles en numerosos idiomas.

d. Evaluación de los resultados.

- Tesauro semiautomático: es una metodología que de una u otra manera combina las anteriores. Lauser *et al.* (2008) concluyen, a partir de una investigación sobre tesauros de agricultura, que los términos provenientes de las ciencias duras como la química y la biología son más adecuados para su extracción por medios automáticos, no así los términos de otras áreas —como la economía o la administración— que requieren mayor trabajo intelectual de los expertos.

Cabe resaltar la utilidad de los estudios bibliométricos como apoyo a cualquiera de las metodologías comentadas. Schneider (2004) señala que la contribución de estos métodos se basa en a) la capacidad del análisis de co-citación y del análisis del contexto de citas para identificar específicamente importantes términos candidatos para el tesauro porque surgen de los contextos y acuerdos de la comunidad científica, y b) la capacidad del análisis de cocitación y de cotérminos para detectar las relaciones semánticas entre los términos.

Autoevaluación

¿Qué opciones pueden elegirse para la creación de la base de datos del tesaurus?

¿Qué capacidades debe tener el programa informático de gestión de tesauros?

¿Qué limitaciones debe tener el programa informático de gestión de tesauros?

¿Qué debe validar el programa informático de gestión de tesauros?

¿Qué vinculaciones debe facilitar el programa informático de gestión de tesauros?

¿Con qué modelos debe ser compatible para funcionar en la Web?

¿Qué es el modelo SKOS?

Desde el punto de vista informático, ¿qué tipo de tesauros hay y en qué consisten?

Bibliografía

- Accomazzi, A., Gray, N., Erdmann, C., Biemesderfer, C., Frey, K. & Soles, J. (2014). The Unified Astronomy Thesaurus. En N. Manset & P. Forshay (Eds.), *ADASS XXIII Proceedings ASP Conference Series. Astronomical Society of the Pacific*. Recuperado de <http://arxiv.org/pdf/1403.6656v1.pdf>
- Asociación Española de Normalización (AENOR) (1990). *Directrices para el establecimiento y desarrollo de tesauros monolingües*. (UNE 50106:1990). Madrid: La Asociación.
- Aguado, A., Pisarello, R., Martínez, A. M. & Delorenci, C. (1999). Definición: nuevas tendencias en tesauros y otros vocabularios controlados. En M. J. López Huertas & J. C. Fernández Molina (Eds), *La representación y la organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información: Actas del IV Congreso ISKO España EOCONSID'99* (pp. 127-132). Granada: Universidad de Granada. Recuperado de <http://www.ugr.es/~isko>
- Aitchison, J. & Dextre Clarke, S. (2004). The thesaurus: a historical viewpoint with a look to the future. En S. Roe & A. R. Thomas (Eds.), *The thesaurus: review, renaissance, and revision* (pp. 5-21). Binghamton NY: Haworth Press.
- Aitchison, J., Gilchrist, A. & Bawden, D. (2000). *Thesaurus construction and use: a practical manual*. (4th ed.). London: Aslib.
- Aitchison, J., Gomersall, A. & Ireland, R. (1969). *Thesaurofacet: a thesaurus and faceted classification for engineering and related subjects*. Whetstone: English Electric Company.
- Almazán, M. B. (2012). *Herramientas para la interoperabilidad y normalización de datos en RI* [Tesis de grado]. Universidad Nacional de La Plata, La Plata. Recuperado de <http://sedici.unlp.edu.ar/handle/10915/25088>

- Álvaro Bermejo, C., Villagr   Rubio,   . & Sorli Rojo,   . (1989a). Desarrollo de lenguajes documentales formalizados en lengua espa  ola: una evaluaci  n: I: vigencia te  rica y pr  ctica de los lenguajes controlados. *Revista Espa  ola de Documentaci  n Cient  fica*, 12(2), 154-159.
-   lvaro Bermejo, C., Villagr   Rubio,   . & Sorli Rojo,   . (1989b). Desarrollo de lenguajes documentales formalizados en lengua espa  ola: II: evaluaci  n de los tesauros disponibles en lengua espa  ola. *Revista Espa  ola de Documentaci  n Cient  fica*; 12(3), 283-297.
- Anderson, J. D. (2003). Organization of knowledge. En: J. Feather & P. Sturges (Eds.), *International encyclopedia of information and library science* (2nd ed., pp.471-490). New York: Routledge.
- Angelozi, S. M. & Mart  n, S. G. (2010). *Metadatos para la descripci  n de recursos electr  nicos en l  nea: an  lisis y comparaci  n de metadatos*. Buenos Aires: Alfagrama.
- Angjeli, A. & Isaac, A. (2008a). Web sem  ntica e interoperabilidad de vocabularios: un experimento en el   mbito de los manuscritos iluminados. En *World Library and Information Congress, IFLA General Conference and Council: Libraries Without Borders: Navigating Towards Global Understanding*. Qu  bec: IFLA. Recuperado de http://archive.ifla.org/IV/ifla74/papers/129-Angjeli_Isaac-trans-es.pdf
- Angjeli, A. & Isaac, A. (2008b). Semantic Web and vocabulary interoperability: an experiment with illumination collections. *International Cataloging and Bibliographic Control*, 38(2), 25-20.
- Arest, A. & Ceretta, M. G. (Comps.) (2001). *Actas del IV Encuentro de Directores y III de Docentes de Escuelas de Bibliotecolog  a y Ciencia de la Informaci  n del Mercosur*. Montevideo: Universidad de la Rep  blica.
- Barit  , M. (2000a). Los conceptos y su representaci  n: una perspectiva terminol  gica para el tratamiento tem  tico de la informaci  n. *Scire*, 6(1), 31-53. Recuperado de <http://ibersid.eu/ojs/index.php/scire/article/view/1123/1105>
- Barit  , M. (2000b). The notion of category: its implications in subject analysis and in the construction and evaluation of indexing languages. *Knowledge Organization*, 27(1/2), 4-10.
- Barit  , M. (2001). Referenciales te  ricos vigentes en el   rea de tratamiento tem  tico de la informaci  n y su expresi  n metodol  gica. En A. Arest &

- M. G. Ceretta (Comps.), *Actas del IV Encuentro de Directores y III de Docentes de Escuelas de Bibliotecología y Ciencia de la Información del Mercosur* (pp. 224-233). Montevideo: EUBCA.
- Barité, M. (2009). Garantía literaria y normas para construcción de vocabularios controlados: aspectos epistemológicos y metodológicos. *Scire*, 15(2), 13-24. Recuperado de <http://ibersid.eu/ojs/index.php/scire/article/view/3709/3470>
- Barité, M. (2011a). *La garantía literaria como herramienta de revisión de sistemas de organización del conocimiento: modelo y aplicación (Tesis doctoral)*. Universidad de Granada, Granada. Recuperado de <http://digibug.ugr.es/bitstream/10481/17583/1/19711864.pdf>
- Barité, M. (2011b). La garantía cultural como justificación en sistemas de organización del conocimiento: aproximación crítica. *Palabra clave* (La Plata), 1(1), 2-11. Recuperado de <http://www.palabraclave.fahce.unlp.edu.ar/article/view/PCv1n1a03/1869>
- Barité, M., Cabrera Castromán, G., Colombo, S., Duarte Blanco, A., Odella, M. L., Simón, L. & Vergara M. (2013). *Diccionario de organización del conocimiento: clasificación, indización, terminología*. (5a ed.). Montevideo: PRODIC. Recuperado de [http://archivos.liccom.edu.uy/diccionario/Diccionario%20Definitivo%20\[3-11-13\].html](http://archivos.liccom.edu.uy/diccionario/Diccionario%20Definitivo%20[3-11-13].html)
- Bates, M. J. (1986). Subject access in online catalogs: a design model. *Journal of the American Society for Information Science*, 37(6), 367-376.
- Bates, M. J. (1988). How to use controlled vocabularies more effectively in online searching. *Online*, 12(6), 45-56.
- Bates, M. J. (1989). Rethinking subject cataloging in online environment. *Library Resources and Technical Services*, 33(4), 400-412.
- Beghtol, C. (1986). Semantic validity: concepts of warrant in bibliographic classification systems. *Library Resources & Technical Services*, 30(2), 109-123.
- Bhattacharyya, G. (1982). Classaurus: its fundamentals, design and use. En I. Dahlberg (Ed.), *Universal classification: subject analysis and ordering systems: Proceedings of the 4th International Study Conference on Classification Research and 6th Annual Conference of Gesellschaft für Klassifikation* (pp.139-148) Frankfurt: Indeks Verlag.
- Bliss, H. E. (1929). *The organization of knowledge and the system of the*

- sciences. New York: H.W. Wilson.
- Bliss, H. E. (1933). *The organization of knowledge in libraries and the subject-approach to books*. New York: H.W. Wilson.
- Boccato, V. (2009). A linguagem documentária vista pelo conteúdo, forma e uso na perspectiva de catalogadores e usuarios. En M. Fujita (Org.), *A indexação de libros: a percepção de catalogadores e usuarios de bibliotecas universitárias: um estudo de observação do contexto sociocognitivo com protocolos verbais*. São Paulo: Cultura Acadêmica. Recuperado de <http://repositorio.unesp.br/handle/11449/109109>
- Bräscher, M. & Café, L. (2008) Organização da Informação ou Organização do Conhecimento?. En *IX ENANCIB: Diversidade Cultural e Políticas de Informação*. São Paulo: Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação. Recuperado de <http://www.slideshare.net/doritchka/brascher-e-caf-organizacao-da-informao-ou-do-conhecimento>
- Broughton, V. (2011). Brian Vickery and the Classification Research Group: the legacy of faceted classification. En *Facets of knowledge organization: a tribute to Professor Brian Vickery 1918-2009: ISKO UK Bienal Conference*. London: ISKO UK. Recuperado de http://www.iskouk.org/sites/default/files/ISKOUK-2011_VandaBroughton.pdf
- Broughton, V., Hansson, J., Hjørland, B. & López Huertas, M. J. (2005). Knowledge organization. En: L. Kajberg & L. Lorrington (Eds.), *European curriculum reflections on library and information science education* (pp. 133-148). Copenhagen: Royal School of Information Science.
- BSI Group (2005-2008). *Structured vocabularies for information retrieval: guide*. (BSI 8723-1-5:2007). London: The Group.
- Business dictionary* (2013). Web Finance. Recuperado de <http://www.businessdictionary.com/>
- Café, L. M. A., Lacruz, M. C. & Barros, C. (2012). Organização do conhecimento: análise conceitual. En M. C. Pérez Pais & M. Á. González Bonome (Eds.), *Actas del X Congreso ISKO-España* (pp. 283-302). La Coruña: Universidad de la Coruña. Recuperado de http://ruc.udc.es/dspace/bitstream/2183/11615/1/CC_132_art_18.pdf
- Centro Argentino de Información Científica y Tecnológica (CAICYT). Grupo de Trabajo sobre Lenguajes de Indización. (1980). *Inventario*

- de lenguajes documentarios no tradicionales en lengua española y portuguesa*. Buenos Aires: El Centro.
- Chan, L. M. & Zeng, L. (2002a). Ensuring interoperability among subject vocabularies and knowledge organization schemes: a methodological analysis. *IFLA journal*, 28(5/6), 323-327.
- Chan, L. M. & Zeng, L. (2002b). Asegurando la interoperabilidad entre vocabularios de materia y esquemas de organización del conocimiento. En *IFLA Council and General Conference*. The Hague: IFLA. Recuperado de <http://archive.ifla.org/IV/ifla68/papers/008-122s.pdf>
- Chaumier, J. (2004). *Les techniques documentaires*. (9e ed.). Paris: Presses Universitaires France.
- Choo, C. (2003). Perspectives on managing knowledge in organizations. En N. J. Williamson & C. Beghtol (Eds.), *Knowledge organization and classification in international information retrieval* (pp. 205-220). New York: Haworth Information Press.
- Chu, H. (2003). *Information representation and retrieval in the digital age*. Medford NJ: American Society for Information Science and Technology (*ASIST monograph series*).
- Claverdon, C. W. & Mills, J. (1985). The testing of index language devices. En L. M. Chan, P. A. Richmond & E. Svenonius (Eds.), *Theory of subject analysis: a sourcebook* (pp.221-246). Littleton: Libraries Unlimited.
- Cleveland, D. B. & Cleveland, A. D. (2013). *Introduction to indexing and abstracting*. (4th ed.). Westport CT: Libraries Unlimited.
- Coates, E. J. (1988). *Subject catalogs: headings and structure*. (2nd ed.). London: Library Association Publishing.
- Coates, E. J., Lloyd, G. & Simandl, D. (1981). *BSO: Systeme general de classement: tables et index*. (3a rev.). París: FID, Unesco.
- Commission des Communautés Européennes (1976). *Définition des caractéristiques essentielles des thesauri: Rapport final* (Vol. 1). Bruxelles: Bureau Marcel van Dijk.
- Cordeiro, M. I. (2003). Knowledge organization from libraries to the Web: strong demands on the weakest side of international librarianship. En N. J. Williamson & C. Beghtol (Eds.), *Knowledge organization and classification in international information retrieval* (pp.65-79). New York: Haworth Information Press.

- Currás, E. (1998). *Tesauros: manual de construcción y uso*. Madrid: Kaher II. Recuperado de http://www.uam.es/personal_pdi/ciencias/ecurras/tesaurosmanualdeconstruccionyuso.pdf
- Dahlberg, I. (1978). Teoria do conceito. *Ciência da Informação*, 7(2), 101-106. Recuperado de <http://revista.ibict.br/ciinf/index.php/ciinf/article/view/1680/1286>
- Dahlberg, I. (1993). Knowledge organization: its scope and possibilities. *Knowledge Organization*, 20(4), 211-222.
- Dahlberg, I. (2006a). Knowledge organization: a new science? *Knowledge organization*, 33(1), 11-19.
- Dahlberg, I. (2006b). Definitionen aus dem Begriffs feld “Wissens organisation”. En: *Internationale Gesellschaft für Wissens organization*, Deutsche Sektion V. Recuperado de <http://www.isko-de.org/index.php?id=dahlbergdefinitionen>
- Dalby, A. (1986). Sumerian catalogs. *The Journal of Library History*, 21(3), 475-487.
- Dublin Core Metadata Initiative (DCMI). (2012). *Dublin Core Metadata Element Set*. Recuperado de <http://dublincore.org/documents/dces/>
- Dextre Clarke, S. G. (2008). The last 50 years of knowledge organization: a journey through my personal archives. *Journal of Information Science*, 34(4), 427-437.
- Dijk, M. & Slipe, G. (1972). *El servicio de documentación frente a la explosión de la información*. Buenos Aires: Consejo Nacional de Investigación Científica y Técnica.
- Doorn, M. & Polman, K. (2010). From classification to thesaurus ... and back?: subject indexing tools at the Library of the Afrika-Studiecentrum Leiden. *Knowledge organization*, 37(3), 203-208.
- Dubois, P. (1987). Free text vs. controlled vocabulary: a reassessment. *Online Review*, 11(4), 243-253.
- Dwyer, J. (1991). Enhancing access to information: designing catalogs for the 21st Century. *Cataloging & Classification Quarterly*, 13(3/4), 29-51.
- Dykstra, M. (1988). LC subject headings disguised as a thesaurus. *Library Journal*, 113(4), 42-46.
- Ensor, P. (1992). User practices in keyword and boolean searching on the online public access catalog. *InformationTechnology and Libraries*,

11(3), 210-219.

- Esteban Navarro, M. A. (1996). El marco disciplinar de los lenguajes documentales: la organización del conocimiento y las ciencias sociales. *Scire*, 2(1), 93-107. Recuperado de <http://www.ibersid.eu/ojs/index.php/scire/article/view/1051>
- Esteban Navarro, M. A. & García Marco, F. J. (1995). Las Primeras Jornadas sobre Organización del Conocimiento: Organización del Conocimiento e Información Científica. *Scire*, 1(1), 149-157. Recuperado de <http://www.ibersid.eu/ojs/index.php/scire/article/view/1038/1020>
- Farradane, J. E. (1952). A scientific theory of classification and indexing: further considerations. *Journal of Documentation*, 8(2), 73-92.
- Ferreira, D. (2009). Explotación de repositorios OAI a través de la armonización de vocabularios controlados. En *Encuentro Nacional de Catalogadores*. Buenos Aires: Biblioteca Nacional. Recuperado de http://www.bn.gov.ar/descargas/catalogadores/ponencias/251109_11a.pdf
- Fidel, R. (1991). Searchers' selection of search keys: III. Searching styles. *Journal of the American Society for Information Science*, 42(7), 515-527.
- Fidel, R. (1992). Who needs controlled vocabulary? *Special Libraries*, 83(1), 1-9.
- Foskett, A. C. (1996). *The subject approach to information*. 5th ed. London: Library Association Publishing.
- Frâncu, V. & Sabo, C. (2010). Implementation of a UDC-based multilingual thesaurus in a library catalogue: the case of BiblioPhil. *Knowledge organization*, 37(3), 209-215.
- Fujita, M. (2008). Organização e representação do conhecimento no Brasil: análise de aspectos conceituais e da produção científica do ENANCIB no período de 2005 a 2007. *Tendências da Pesquisa Brasileira em Ciência da Informação*, 1(1). Recuperado de <http://inseer.ibict.br/ancib/index.php/tpbci/article/view/4>
- García de Kohen, M. & Monfasani, R. E. (1984). Tesauro argentino de psicología. *Revista Española de Documentación Científica*, 7(3), 229-232.
- García de Kohen, M. & Monfasani, R. E. (2009). *Tesauro de psicología*. Buenos Aires: Dirección de Biblioteca de la Facultad de Psicología de la Universidad de Buenos Aires. Recuperado de <http://bvs.psi.uba.ar/local/File/tesauro/TESAUROpresent.pdf>

- García Gutiérrez, A. L. (1990). *Estructura lingüística de la documentación: teoría y método*. Murcia: Universidad de Murcia.
- García Marco, F. J. (1995). Los contenidos y la secuencia docente de la organización y representación del conocimiento: una propuesta interdisciplinar. *Organización del Conocimiento en Sistemas de Información y Documentación*, 1, 219-228. Recuperado de <http://dialnet.unirioja.es/descarga/articulo/2341341.pdf>
- García Marco, F. J. (2006). Los múltiples afloramientos de la organización del conocimiento en el ciclo de la información documental. *Scire*, 12(1), 9-22. Recuperado de <http://www.ibersid.eu/ojs/index.php/scire/article/view/1583>
- Gil Leiva, I. (2008). *Manual de indización: teoría y práctica*. Gijón: Trea.
- Gil Leiva, I. & Rodríguez M., J. (1997). Análisis de los descriptores de diferentes áreas de conocimiento indizadas en bases de datos del CSIC: Aplicación a la indización automática. *Revista Española de Documentación Científica*, 20(2). Recuperado de <http://redc.revistas.csic.es/index.php/redc/article/viewArticle/589>
- Gil Urdiciain, B. (2004). *Manual de lenguajes documentales*. (2a ed.). Gijón: Trea.
- Gilchrist, A. (2011). A note on the theory and practice of knowledge organization and knowledge representation. *Scire*, 17(1), 23-28. Recuperado de <http://www.ibersid.eu/ojs/index.php/scire/article/view/3920/3635>
- Gnoli, C. (2007). Progress in synthetic classification: towards unique definitions of concepts. En *UDC Seminar*. The Hague: UDC Consortium. Recuperado de <http://arizona.openrepository.com/arizona/bitstream/10150/105614/1/synthetic.pdf>
- Gnoli, C. (2011). Ontological foundations in knowledge organization: the theory of integrative levels applied in citation order. *Scire*, 17(1), 29-34.
- Gottschalk, P. (2007). *Knowledge management systems: value shop creation*. Hershey PA: Idea Group.
- Granados Colillas, M. & Nicolau, A. (2007). Aplicación de un nuevo sistema de indización en una colección de recursos especializados en ciencias de la educación. *El profesional de la información*, 16(6), 627-635. Recuperado de <http://www.elprofesionaldelainformacion.com/>

[contenidos/2007/noviembre/09.pdf](#)

- Granados Colillas, M. & Nicolau, A. (2008). *Improving subject searching in databases through a combination of descriptors and UDC*. En *BOBCATSSS'08: Providing acces for everyone*. Zadar. Recuperado de <http://eprints.rclis.org/11661/>
- Guimarães, J. A. (2001a). *Ensino de tratamento temático da informação nos cursos de biblioteconomía do Mercosur: análise e perspectivas de un core curriculum à luz dos avanços teóricos da área de organização do conhecimento*. Marília: Faculdade de Filosofia e Ciências da UNESP.
- Guimarães, J. A. (2001b). Brasil: principales tendencias y enfoques en el área. En A. Arest & M. Ceretta (Comps.), *Actas del IV Encuentro de Directores y III de Docentes de Escuelas de Bibliotecología y Ciencia de la Información del Mercosur* (pp. 336-339). Montevideo: Universidad de la República.
- Guimarães, J. A. (2008). A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). *Revista Ibero-americana de Ciência da Informação*, 1(1), 77-99. Recuperado de <http://seer.bce.unb.br/index.php/RICI/article/view/2761/2331>
- Guimarães, J. A. (2009). Abordagens teóricas de tratamento temático da informação (TTI): catalogação de assunto, indexação e análise documental. *Ibersid: Revista de Sistemas de Información y Documentación*, 3, 107-117. Recuperado de <http://www.ibersid.eu/ojs/index.php/ibersid/article/view/3730>
- Guimarães, J. A., Pinho, F. & Ferreira, G. (2012). Relações teóricas da organização do conhecimento com as abordagens de catalogação de assunto, indexação e análise documental: uma análise de domínio da revista Scire (1995-2010). *Scire*, 18(2), 31-41. Recuperado de <http://www.ibersid.eu/ojs/index.php/scire/article/view/3971>
- Guinchat, C. & Menou, M. (1990). *Introducción general a las ciencias y técnicas de la información y la documentación*. (2a ed.). París: Unesco.
- Hill, L. & Koch, T. (2001). Networked knowledge organization systems: introduction to a special issue [editorial]. *Journal of Digital Information*, 1(8). Recuperado de <http://journals.tdl.org/jodi/article/view/32/33>
- Hjerpe, R. (1990). A framework for characterizing systems for knowledge

- organization: a first basis for comparisons and evaluations. *Advances in Knowledge Organization*, 3, 21-34.
- Hjerpe, R. (1996). Go with the flow, or abide by the side, or watch the waves?: challenges of change for knowledge organization. *Advances in Knowledge Organization*, 5, 10-25.
- Hjørland, B. (2006). Indexing language. En *Lifeboat for knowledge organization*. Recuperado de http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/indexing_languages.htm
- Hjørland, B. (2007a). Semantics and knowledge organization. *Annual Review of Information Science and Technology*, 41, 367-404.
- Hjørland, B. (2007b). Semantic tool. En *Lifeboat for knowledge organization*. Recuperado de http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/semantic_tools.htm
- Hjørland, B. (2008a). What is knowledge organization (KO)? *Knowledge Organization*, 35, (2/3), 86-102.
- Hjørland, B. (2008b). Knowledge organization systems. En *Lifeboat for knowledge organization*. Recuperado de http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/knowledge_organization_systems.htm
- Hjørland, B. & Albrechtsen, H. (1995). Toward a new horizon in information science: domain-analysis. *Journal of the American Society for Information Science*, 46(6), 400-425.
- Hodge, G. (2000). *Systems of knowledge organization for digital libraries: beyond traditional authority files*. Washington DC: Digital Library Federation, Council on Library and Information Resources. Recuperado de <http://clir.org/pubs/reports/pub91/contents.html>
- Hudon, M. (1996). Preparing terminological definition for indexing and retrieval thesauri: a model. *Advances in Knowledge Organization*, 5, 363-369.
- Hudon, M. (2003). True and tested products: thesauri on the Web. *The Indexer*, 23(3), 115-119.
- Instituto Colombiano de Normas Técnicas y Certificación (1998). *Documentación: Guía para el establecimiento y desarrollo de tesauros monolingües*. (2a actualización). (NTC 1476:1998). Bogotá: El Instituto.
- International Federation of Library Associations and Institutions (IFLA) (2004). *Requisitos funcionales de los registros bibliográficos*. Madrid:

- Ministerio de Cultura. Recuperado de <http://www.ifla.org/files/assets/cataloguing/frbr/frbr-es.pdf>
- International Federation of Library Associations and Institutions (IFLA) (2009a). *Guidelines for multilingual thesauri* (IFLA professional reports, No 115). Recuperado de <http://archive.ifla.org/VII/s29/pubs/Draft-multilingualthesauri.pdf>
- International Federation of Library Associations and Institutions (IFLA) (2009b). *Requisitos funcionales de los datos de autoridad (FRAD): un modelo conceptual*. Recuperado de http://www.ifla.org/files/assets/cataloguing/frad/frad_2009-es.pdf
- International Federation of Library Associations and Institutions (IFLA) (2010). *Functional requirements for subject authority data: a conceptual model*. Recuperado de <http://www.ifla.org/files/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf>
- International Federation of Library Associations and Institutions (IFLA) (2011). *Guidelines for subject access in national bibliographies*. Recuperado de http://www.ifla.org/files/assets/classification-and-indexing/subject-access-by-national-bibliographic-agencies/nba_guidelines_draft_2011-05.pdf
- Instituto Argentino de Normalización y Certificación (IRAM) (1983). *Tesauros monolingües para la recuperación de información: estructura y desarrollo*. (IRAM 32057:1983). Buenos Aires: El Instituto.
- Instituto Argentino de Normalización y Certificación (IRAM) (2007). *Información y documentación: indicadores de desempeño en bibliotecas*. (IRAM-ISO 11620:2007). Buenos Aires: El Instituto.
- International Society for Knowledge Organization (ISKO). (1996) *Compatibility and integration of order systems: research seminar proceedings of the TIP/ISKO Meeting*. Warsaw: Stowarzyszenie Bibliotekarzy Polskich.
- International Organization for Standardization (ISO) (1985). *Documentation: guidelines for the establishment and development of multilingual thesauri*. (ISO 5964:1985). Geneva: The Organization.
- International Organization for Standardization (ISO) (1986). *Documentation: guidelines for the establishment and development of monolingual*

- thesauri*. (ISO 2788:1986). (2nd ed.). Geneva: The Organization.
- International Organization for Standardization (ISO) (2001). *Information and documentation: vocabulary*. (ISO 5127:2001). Geneva: The Organization.
- International Organization for Standardization (ISO) (2002). *Codes for the representation of names of languages. Part 1: Alpha-2 code*. (ISO 639-1:2002). Geneva: The Organization.
- International Organization for Standardization (ISO) (2009). *Terminology work: principles and methods = Travail terminologique: principes et methodes*. (ISO 704:2009). Geneva: The Organization.
- International Organization for Standardization (ISO) (2011). *Information and documentation: thesauri and interoperability with other vocabularies: Part 1. Thesauri for information retrieval*. (ISO 25964-1:2011). Geneva: The Organization.
- International Organization for Standardization (ISO) (2013). *Information and documentation: thesauri and interoperability with other vocabularies: Part 2. Interoperability with other vocabularies*. (ISO 25964-2:2013). Geneva: The Organization .
- International Standard Serial Number International Center (ISSN). (2011). *Total number of records in the ISSN register*. Recuperado de <http://www.issn.org/files/issn/statistiques/Total-number-of-records.pdf>
- Kilgour, F. G. (1998). Origins of coordinate searching. En T. B. Hahn & M. K. Buckland (Eds.), *Historical studies in information science* (pp. 107-115). Medford NJ: Information Today.
- Kless, D., Jansen, L., Lindenthal, J. & Wiebensohn, J. (2012). A method for re-engineering a thesaurus into an ontology. En M. Donnelly & G. Guizzardi (Eds.), *ACTAS DE LA SÉPTIMA CONFERENCIA INTERNACIONAL Formal Ontology and Information Systems* (pp. 133-146). [s.l.]: **IOS PRESS**
- Lancaster, F. W. (1985). *A thesaurus construction and use: a condensed course*. (PGI-85/WS/11). Paris: Unesco. Recuperado de <http://unesdoc.unesco.org/images/0007/000703/070359eb.pdf>
- Lancaster, F. W. (1995). *Indización y resúmenes: teoría y práctica*. Buenos Aires: EB Publicaciones.
- Lancaster, F. W. (2002) [1986]. *El control del vocabulario en la recuperación de información*. (2a ed.). Valencia: Universitat de Valencia (*Colleccio*

Educació. Materials).

- Lauser, B., Johannsen, G., Caracciolo, C., Hage, W. R., Keizer, J. & Mayr, P. (2008). *Comparing human and automatic thesaurus mapping approaches in the agricultural domain*. En *International Conference on Dublin Core and Metadata Applications*. Berlin. Recuperado de http://www.ib.hu-berlin.de/~mayr/arbeiten/DC2008_Lauser-et al.pdf
- Liang, A. C., Sini, M., Chun, Ch., Sijing, L., Wenlin, L., Chunpei, H. & Keizer, J. (2005). The mapping schema from Chinese Agricultural Thesaurus to AGROVOC. En *Agricultural Ontology Service Workshop on Ontologies: the More Practical Issues and Experience*. Rome: Food and Agriculture Organization. Recuperado de <http://www.fao.org/3/a-af241e.pdf>
- López Yepes, J. (Coord.) (2004). *Diccionario enciclopédico de ciencias de la documentación*. Madrid: Síntesis.
- Lyons, J. (1984). *Introducción al lenguaje y a la lingüística*. Barcelona: Teide.
- Lyons, J. (1997). *Semántica lingüística: una introducción*. Barcelona: Paidós Ibérica.
- Ma, X., Carranza, E. J.M., Wu, C., Meer, F. D. & Liu, G. (2011). A SKOS-based multilingual thesaurus of geological time scale for interoperability of online geological maps. *Computers & Geosciences*, 37, 1602-1615. Recuperado de http://tw.rpi.edu/media/2013/03/07/a10a/skos_cg.pdf
- Macgregor, G., Joseph, A. & Dennis, N. (2007). A SKOS core approach to implementing and M2M terminology mapping server. En: *International Conference on Semantic Web and Digital Libraries*. Bangalore. Recuperado de <http://strathprints.strath.ac.uk/2970/1/strathprints002970.pdf>
- Martínez Tamayo, A. M. & Mendes, P. V. (2010). La norma argentina de construcción de tesauros IRAM 32057: necesidad de actualización. 1. El vocabulario controlado. *Información, cultura y sociedad*, 21, 11-30. Recuperado de <http://www.scielo.org.ar/pdf/ics/n21/n21a02.pdf>
- Martínez Tamayo, A. M. & Valdez, J. C. (2009). *Indización y clasificación en bibliotecas*. Buenos Aires: Alfagrama.
- Martínez Tamayo, A., Ristuccia, C. A., Stubbs, E. A., Valdez, J. C., Gamba, V. L., Mendes, P. V., Unzurrunzaga, C. & Caminotti, M. (2009). Concepto, forma y longitud de los términos preferentes del tesoro: una propuesta de indicadores de calidad. *Anales de Documentación*, 13, 185-195. Recuperado de <http://revistas.um.es/analesdoc/article/>

[viewFile/107151/101821](#)

- Martínez Tamayo, A., Ristuccia, C. A., Stubbs, E. A., Valdez, J. C., Gamba, V. L., Mendes, P. V., Unzurrunzaga, C. & Caminotti, M. (2011a). La estructura sistemática del tesaurus: indicadores para evaluar su calidad. *Revista Española de Documentación Científica*, 34(1), 29-43. Recuperado de <http://redc.revistas.csic.es/index.php/redc/article/view/681/755>.
- Martínez Tamayo, A. M., Valdez, J. C., Stubbs, E. A., González Terán, Y. & Kesler, M. I. (2011b). Interoperabilidad de sistemas de organización del conocimiento: el estado del arte. *Información, Cultura y Sociedad*, 24, 15-37. Recuperado de <http://www.scielo.org.ar/pdf/ics/n24/n24a02.pdf>
- Mayr, P. & Petras, V. (2008). Cross-concordances: terminology mapping and its effectiveness for information retrieval. En *World Library and Information Congress, IFLA General Conference and Council*. The Hague: IFLA. Recuperado de http://archive.ifla.org/IV/ifla74/papers/129-Mayr_Petras-en.pdf
- Mayr, P. & Petras, V. (2009). Cross-concordances: terminology mapping and its effectiveness for information retrieval. En *International cataloging and bibliographic control*, 38(3), 43-52.
- McCulloch, E. & Macgregor, G. (2008). Analysis of equivalence mapping for terminology services. En *Journal of information science*, 34(1), 70-92.
- Mijailov, A. I. & Guiliarevskii, R. S. (1974). *Curso introductorio de informática/documentación*. Caracas: Fundación Instituto Venezolano de Productividad.
- Miller, J. & Bristow, B. A., (Eds.) (2007). *Sears list of subject headings*. (19th ed.). New York: H.W. Wilson.
- Miller, U. (1997). Thesaurus construction: problems and their roots. *Information Processing and Management*, 33(4), 481-493.
- Miller, U. & Teitelbaum, R. (2002). Pre-coordination and post-coordination: past and future. *Knowledge Organization*, 29(2), 87-93.
- Mills, J. (1960). *A modern outline of library classification*. (2nd ed.). London: Chapman & Hall.
- Mills, J. (2004). Library classification and logical division in information retrieval. *Library Trends*, 52(3), 541-570.
- Molina, C. & Zítara, E. (1998). Actividades terminológicas en la Argentina: enfoque desde la bibliotecología y la documentación. En *Actas del VI*

Simposio de RITERM. La Habana: RITERM.

- Naumis Peña, C. (2003). Indización y clasificación: un problema conceptual y terminológico. *Documentación de las Ciencias de la Información*, 26, 23-40. Recuperado de <http://revistas.ucm.es/index.php/DCIN/article/view/DCIN0303110023A/19329>
- Naumis Peña, C. (2008). *Los tesauros documentales y su aplicación en la información impresa, digital y multimedia*. Buenos Aires: Alfagrama.
- Neelamegan, A. & Lalitha, S. K. (2013). Multilingual thesaurus and interoperability. *DESIDOC Journal of Library & Information Technology*, 33(4), 289-294. Recuperado de <http://www.publications.drdo.gov.in/ojs/index.php/djlit/article/view/4883/2841>
- Nicholson, D. (2002). Subject-based interoperability: issues from the high level thesaurus (HILT) Project. En *IFLA Council and General Conference*. Glasgow. Recuperado de <http://archive.ifla.org/IV/ifla68/papers/006-122e.pdf>
- Nicholson, D. & McCulloch, E. (2006). Interoperable subject retrieval in a distributed multi-scheme environment: new developments in the HILT Project. En *Scire*, 12(1), 109-124. Recuperado de <http://ibersid.eu/ojs/index.php/scire/article/view/1589/1561>
- National Information Standards Organization (NISO) (2010). *Guidelines for the construction, format, and management of monolingual controlled vocabularies*. (ANSI/NISO Z39.19:2005 [R2010]). Bethesda MD: NISO Press. Recuperado de http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf
- Oficina Nacional de Normalización (2000). *Directrices para el establecimiento y desarrollo de tesauros monolingües*. (Norma cubana 2788:2000). La Habana: La Oficina. Recuperado de http://www.sld.cu/galerias/pdf/sitios/centromed/nc_iso_202788_directrices_para_el_establecimiento_y_desarrollo_de_tesauros_monolingues.pdf
- Owens, L. A. & Cochrane, P. A. (2004). Thesaurus evaluation. En S. K. Roe & A. R. Thomas (Eds.), *The thesaurus: review, renaissance, and revision* (pp. 87-102). New York: Haworth Information Press.
- Peña Vera, T. (2010). El acceso al saber requiere aplicar organización y representación del conocimiento. *Revista Interamericana de Bibliotecología*, 33(1), 13-29. Recuperado de <http://eprints.rclis.>

[org/17281/1/5800.pdf](http://www.elsevier.com/locate/S0022-3778(03)00050-1)

- Peña Vera, T. (2011). *Organización y representación del conocimiento: incidencia de las tecnologías de la información y la comunicación*. Buenos Aires: Alfagrama.
- Pinto Molina, M., (Ed.) (2001). *Catalogación de documentos: teoría y práctica*. (2a ed.). Madrid: Síntesis (*Bibliotecología y Documentación*).
- Ranganathan, S. R. (1933). *Colon classification*. Madras: The Madras Library Association.
- Ranganathan, S. R. (1937). *Prolegomena to library classification*. Madras: The Madras Library Association.
- Ranganathan, S. R. (1967). *Prolegomena to library classification*. (3rd ed.). London: Asian Publishing House. Recuperado de <http://arizona.openrepository.com/arizona/handle/10150/106370>
- Real Academia Española (2013). *Diccionario de la lengua española*. (22a ed.). Madrid: La Academia. Recuperado de <http://www.rae.es>
- Reverté, C. & Sebastián, M. (2009). Aquatic science subject Gateway Project as a model of interoperability. En *ISKO-UK 2009 Conference*. London: ISKO-UK Chapter.
- Riaz, M. (1989). *Advanced indexing and abstracting practices*. New Delhi: Atlantic Publishers & Distributors.
- Roberts, N. (1984). The pre-history of the information retrieval thesaurus. *Journal of Documentation*, 40(4), 271-285.
- Rolland-Thomas, P. (1989). Thesaural codes: an appraisal of their use in the Library of Congress subject headings. *Cataloging and Classification Quarterly*, 16(2), 71-91.
- Rowley, J. (1994). The controlled versus natural indexing languages debate revisited: a perspective on information retrieval practice and research. *Journal of Information Science*, 20(2), 108-118.
- Ruiz Pérez, R. (1992). *El análisis documental*. Granada: Universidad de Granada.
- Schneider, J. W. (2004). *Verification of bibliometric methods' applicability for thesaurus construction [Tesis de doctorado]*. Copenhagen: Royal School of Library and Information Science.
- Shearer, J. R. (2004). A practical exercise in building thesaurus. En S. K. Roe & A. R. Thomas (Eds.), *The thesaurus: review, renaissance, and revision*

- (35-56). Binghamton NY: Haworth Information Press.
- Shiri, A. & Revie, C. (2006). Query expansion behavior within a thesaurus-enhanced search environment: a user-centered evaluation. *Journal of the American Society for Information Science and Technology*, 57(4), 462-478.
- Souza, R., Tudhope, D. & Almeida, M. (2010). The KOS spectra: a tentative typology of knowledge organization systems. En C. Gnoli & F. Mazzocchi (Eds.), *Paradigms and conceptual systems in knowledge organization: Proceedings of the Eleventh International ISKO Conference* (pp. 122-128). ISKO (*Advances in Knowledge Organization*; 12).
- Souza, R., Tudhope, D. & Almeida, M. (2012). Toward a taxonomy of KOS. *Knowledge Organization*, 39(3), 179-192.
- Spiteri, L. (1998). A simplified model for facet analysis. *Canadian Journal of Information and Library Science*, 23(1), 1-30. Recuperado de http://iainstitute.org/en/learn/research/a_simplified_model_for_facet_analysis.php
- Svenonius, E. (1986). Unanswered questions in the design of controlled vocabularies. *Journal of the American Society for Information Science*, 37(5), 331-340.
- Svenonius, E. (2000). *The intellectual foundation of information organization*. Cambridge MT: MIT Press (*Digital libraries and electronic publishing*).
- Svenonius, E. (2003). Design of controlled vocabularies. En *Encyclopedia of library and information science*. New York: Marcel Dekker. Recuperado de http://polaris.gseis.ucla.edu/gleazer/260_readings/Svenonius.pdf.
- Taylor, A. G. (2004). *The organization of information*. 2nd ed. Westport: Libraries Unlimited.
- Tennis, J. T. (2003). Two axes of domains for domain analysis. *Knowledge Organization*, 30(3/4), 191-195.
- Tennis, J. T. (2004). Layers of meaning: disentangling subject access interoperability. *Advances in classification research*, 12, 113-122. Recuperado de http://faculty.washington.edu/jtennis/Tennis_layers_2001.pdf
- Tenopir, C. (1985). Searching Harvard Business Review Online: lessons in searching a full text database. *Online*, 9, 71-78.
- Todaro, A. J. & Martínez Tamayo, A. M. (2006a). Las razas en una lista de encabezamientos de materia en español. *Ciência da Informação*, 35(3),

- 272-281 Recuperado de <http://www.scielo.br/pdf/ci/v35n3/v35n3a27.pdf>
- Todaro, A. J. & Martínez Tamayo, A. M. (2006b). La mujer en una lista de encabezamientos de materia para bibliotecas. *Investigación Bibliotecológica* (México), 20(41), 195-206. Recuperado de http://www.scielo.org.mx/scielo.php?pid=S0187-358X2006000200009&script=sci_arttext
- Tudhope, D. & Koch, T. (2004). New Applications of Knowledge Organization Systems: introduction to a special issue. *Journal Of Digital Information*, 4(4). Recuperado de <https://journals.tdl.org/jodi/index.php/jodi/article/view/109/108>
- Tudhope, D., Alani, H. & Jones, C. (2001). Augmenting thesaurus relationships: possibilities for retrieval. *Journal of Digital Information*, 1(8). Recuperado de <http://journals.tdl.org/jodi/index.php/jodi/article/view/181>
- Tudhope, D., Koch, T. & Heere, R. (2006). *Terminology services and technology: JISC state of the art review*. JISC. Recuperado de <http://opus.bath.ac.uk/23563/>
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (Unesco) (1971). *Guidelines for the establishment and development of monolingual thesauri for information retrieval*. Paris: La Organización. Recuperado de <http://unesdoc.unesco.org/images/0000/000059/005951eb.pdf>
- Varlejs, J. (2004). The technical report and its impact on post-World War II information systems: two case studies. En W. B. Rayward & M. E. Bowden (Eds.), *The history and heritage of scientific and technological information systems: proceedings of the 2002 Conference* (pp.89-99). Recuperado de https://books.google.com.ar/s?id=76OOQannpBgC&dq=WII+information+kilgour&hl=es&source=gbp_navlinks_s
- Vickery, B. C. (1960a). *Faceted classification: a guide to construction and use of special schemes*. London: Aslib.
- Vickery, B. C. (1960b). Thesaurus: a new word in documentation. *Journal of Documentation*, 16(4), 181-189.
- Vizcaya Alonso, D. (1997). *Lenguajes documentarios*. Rosario: Nuevo Paradigma.
- Vizine Goetz, D., Hickey, C., Houghton, A. & Thompson, R. (2004). Vocabulary mapping for terminology services. *Journal of digital*

- information, 4(4). Recuperado de <http://journals.tdl.org/jodi/article/view/114/113>
- World Wide Web Consortium (W3C) (2009). *Manual de SKOS (simple knowledge organization system): nota del Grupo de Trabajo del W3C*. Recuperado de <http://skos.um.es/TR/skos-primer/>
- World Wide Web Consortium (W3C) (2011). *W3C semantic Web activity*. Recuperado de <http://www.w3.org/2001/sw/>
- Wang, J. (2003). A knowledge network constructed by integrating classification, thesaurus, and metadata in digital library. *International information & library research* 35(2/4), 383-397.
- Wersig, G. & Neveling, U. (1976). *Terminology of documentation*. Paris: Unesco.
- Willer, M., (Ed.). (2009). *UNIMARC manual authorities format*. (3rd ed.). Munich: K.G. Saur. (*IFLA Series on bibliographic control*; 38).
- Wilson, T. D. (2002). The nonsense of knowledge management. *Information Research*, 8(1). Recuperado de <http://informationr.net/ir/8-1/paper144.html>
- Zeng, M. L. (2008). Knowledge organization systems (KOS). *Knowledge Organization*, 35(2/3), 160-182.
- Zeng, M. L. & Zummer, M. (2009a). Mapping FRISAD model and other abstract models. En *ISKO-UK 2009 Conference*. London: ISKO-UK Chapter. Recuperado de <http://www.iskouk.org/content/mapping-frisad-model-and-other-abstract-models>
- Zeng, M. L. & Zummer, M. (2009b). Introducing FRISAD and mapping it with SKOS and other models. En *Proceedings of the World Library and Information Congress, IFLA General Conference and Council*. The Hague: IFLA. Recuperado de <http://conference.ifla.org/past/2009/200-zeng-en.pdf>
- Zeng, W. (2012). Exploration and study of multilingual thesauri automation construction for digital libraries in China. *Electronic Library*, 30(2), 233-247.
- Zhang, X. (2006). Concept integration of document databases using different indexing languages. *Information Processing & Management*, 42(1), 121-135.
- Zítara, E., (Coord.). (1983). *Tesaurus sobre recursos hídricos*. Ezeiza: Instituto Nacional del Agua y del Ambiente.

Las Autoras

Ana María Martínez Tamayo

Licenciada en Bibliotecología y Documentación. UNLP. FaHCE. Especialista en Informática Aplicada a la Educación. UNLP. Facultad de Informática. Se desempeñó como Profesora Titular de Organización del Conocimiento I y II. UNLP. FaHCE (hasta su jubilación en 2014).

Publicaciones

Libro: Martínez, A. & Valdez, J. (2009). *Indización y Clasificación en bibliotecas*. Buenos Aires: Alfagrama.

Capítulo: Martínez, A. (2011). “Evolución del catálogo en línea”. En: PICO, P., coord. *Manual de catalogación automatizada*. Montevideo: Universidad de la República. Escuela Universitaria de Bibliotecología y Ciencias Afines.

Artículo: Martínez, A. & Mendes, P. (2009). “La norma argentina de construcción de tesauros IRAM 32057: necesidad de actualización. 1. El vocabulario controlado”. *Información, Cultura y Sociedad*, 21.

Actividades anteriores

Ha dictado cursos y dirigido tesis de licenciatura y proyectos de investigación sobre temas relacionados con las listas de encabezamiento de materia, sistemas de clasificación, listas de autoridades de materia, tesauros, taxonomías, calidad de los catálogos en línea e interoperabilidad entre los distintos sistemas de organización del conocimiento, que han generado diversos artículos publicados en revistas especializadas.

Paola Verónica Mendes

Licenciada en Bibliotecología y Ciencia de la Información. UNLP.

FaHCE. Profesora ayudante en la cátedra de Organización del Conocimiento II. UNLP. FaHCE. Profesional Bibliotecaria del área de Procesos técnicos de la Biblioteca de Arquitectura. UNLP. Integrante del Proyecto de Investigación sobre Procesamiento de Cartas Náuticas. Miembro del Subcomité de Documentación de IRAM.

Publicaciones

Artículos: Martínez, Ristuccia, Valdez, Gamba, Mendes, Unzurrungaza & Caminotti. (2011). “La estructura sistemática del tesoro: indicadores para evaluar su calidad”. Revista española de documentación científica, 34.

Martínez, Ristuccia, Valdez, Gamba, Mendes, Unzurrungaza & Caminotti. (2010). “Concepto, forma y longitud de los términos preferentes del tesoro: una propuesta de indicadores de calidad”. Anales de Documentación, 13.

Martínez & Mendes. (2009). “La norma argentina de construcción de tesauros IRAM 32057: necesidad de actualización. 1. El vocabulario controlado”. Información, Cultura y Sociedad, 21.

Actividades anteriores

Presentación de ponencias relacionadas con tesauros. Colaboradora en Proyecto de investigación sobre Evaluación de Tesauros. Docente de Análisis documental. ISFDyT nº 8

Índice de figuras y tablas

Los sistemas y procesos de organización del conocimiento	
Tabla 1. Características de los SOC según sus funciones	35
Figura 1. Sistemas de organización del conocimiento (SOC)	37
Tabla 2. Lengua natural vs. SOC	41
Distintos enfoques para estudiar los SOC	
Figura 1. Triángulo del concepto	55
Tabla 1. Entidades del modelo FRBR-FRAD-FRSAD	61
Figura 2. Relación obra-thema	62
Figura 3. Relación thema-nomen	62
El tesoro	
Figura 1. Modelo de los procesos de indización y recuperación	69
Figura 2. Registro bibliográfico de un catálogo en línea	69
Figura 3. Registro metadatos de un repositorio institucional	70
La historia del tesoro	
Figura 1. Crecimiento de las publicaciones periódicas 1800-2000	76
Figura 2. Ejemplo de fichas peek-a-boo de William E. Batten	77
Figura 3. Ejemplo de una ficha Zatocard de Calvin Mooers	79
Figura 4. Ejemplo de fichas Uniterm de Mortimer Taube	80
Diseño y desarrollo del tesoro	
Figura 1. Curva de crecimiento de términos	92
El vocabulario del tesoro	
Figura 1. Vocabularios de indización y de entrada	99

Las relaciones jerárquicas y asociativas	
Figura 1. Prueba de la relación jerárquica género-especie	125
Figura 2. Ejemplo de una relación polijerárquica	128
El orden sistemático del tesoro	
Figura 1. Dominio, categorías y facetas	141
Figura 2. Faceta, características de división y subfacetas	142
Figura 3. Subfacetas de segundo, tercero y cuarto nivel	144
Figura 4. Cadena	145
Tabla 1. Comparación de una notación ordinal y una jerárquica	151
Evaluación, publicación y actualización del tesoro	
Tabla 1. Métodos de evaluación de tesauros	153
Figura 1. Ejemplo de la sección sistemática del tesoro	162
Figura 2. Orden sistemático en forma de mapa de tópicos	163
Figura 3. Ejemplo de un registro individual de término preferente	164
Figura B1. Interfaz de búsqueda del tesoro Agrovoc	167
Figura B2. Lista alfabética de términos del tesoro Agrovoc	168
Figura B3. Registro individual de un término preferente del tesoro Agrovoc	169
Figura B4. Despliegue del orden sistemático del tesoro Agrovoc	170
Figura B5. Interfaz de búsqueda del tesoro Spines	171
Figura B6. Sección sistemática del tesoro Spines	172
Figura B7. Sección alfabética del tesoro Spines	173
Figura B8. Registro individual de un término preferente del tesoro Spines	174
Figura B9. Interfaz del Tesoro de OIT	175
Figura B10. Registro individual de un término preferente del Tesoro de OIT	176
Figura B11. Interfaz del Tesoro Europeo de Educación	177
Figura B12. Registro individual de un término preferente del Tesoro Europeo de Educación	178
El tesoro y la interoperabilidad	
Figura 1. Modelo de interoperabilidad de vinculación directa	181

<u>Figura 2. Modelo de interoperabilidad de estructura de hub</u>	181
<u>Figura 3. Modelo de interoperabilidad de mapeo selectivo</u>	182
Software de gestión, modelo de datos y formatos de intercambio	
<u>Figura 1. Modelo de concepto SKOS tomado del Tesauro de la Unesco</u>	197
<u>Tabla 1. Vocabulario del modelo SKOS</u>	198

La presente obra, de carácter introductorio, está dirigida a estudiantes y profesionales de Bibliotecología, Documentación y Ciencia de la Información de lengua española, que se enfrentan por primera vez al diseño y desarrollo de tesauros. Se encuadra en el marco referencial de la Organización del Conocimiento y se basa principalmente en la nueva norma ISO 25964-1/2 que ha puesto énfasis en la publicación electrónica del tesauro. Sus doce capítulos tratan sobre la Organización del Conocimiento, sus sistemas y los distintos enfoques con que éstos han sido estudiados. Se define y caracteriza el tesauro, se resume su historia, se describe su vocabulario, las relaciones y el orden sistemático, así como su evaluación, publicación y actualización. También se hace referencia a la interoperabilidad entre tesauros y otros sistemas de organización del conocimiento, el software de gestión, los modelos de datos y los formatos de intercambio.

ISBN 978-950-34-1307-4